The Measurement of the Hadronic Energy Flow and Jet Production with the ZEUS Calorimeter in Deep Inelastic Scattering Events at HERA

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
Prof. Dr. P. W. M. de Meijer
in het openbaar te verdedigen in de Aula der Universiteit
(Oude Lutherse Kerk, ingang Singel 411, hoek Spui)
op vrijdag 3 september 1993 te 15.00 uur

door

Paulus Johannes de Jong
geboren te Zwolle

Promotor: Prof. Dr. A. G. Tenner

Faculteit der Natuur- en Sterrenkunde

The work described in this thesis is part of the research programme of 'het Nationaal Instituut voor Kernfysica en Hoge–Energie Fysica (NIKHEF–H)' in Amsterdam, the Netherlands. The author was financially supported by 'de Stichting voor Fundamenteel Onderzoek der Materie (FOM)'.

'Scire est mensurare'
– Johannes Kepler

'It is not poetry, but prose run mad'
– Alexander Pope

Contents

Chapter 1

Introduction

In high energy physics, or elementary particle physics, we are interested in the study of the fundamental building blocks of matter, and, even more important, the interactions between them that enable the existence of stable forms of matter, such as money, fast cars, alcohol, beaches, and women. Experimentally, the study of elementary particles involves looking at very small structures, e.g. distances smaller than the size of a proton, or 10^{-13} cm, which is, ironically, currently only possible with some of the largest scientific instruments that exist. These machines accelerate elementary particles to energies of more than 10^9 eV and let them collide; the debris of the collisions are measured by large and complicated detectors, which serve as the eyes of the elementary particle physicist.

In the past two decades, a collection of physical theories known as the 'standard model' of elementary particle physics has proven to be extremely effective in its description of elementary particles and their interactions in the energy range accessible to present day experiments. The standard model describes the interactions of the fundamental building blocks of matter as we currently know them, the fermions called *quarks* and *leptons*, via three forces: the electromagnetic force acts between electrically charged particles and governs the structure of atoms, the strong nuclear force acts only upon *hadrons* and is responsible for the stability of atomic nuclei and particles within the nucleus, and the weak nuclear force, that acts upon all particles and is often responsible for their transformation and decay. The major success of the standard model is the discovery that, in the current universe, electromagnetic and weak nuclear forces are nothing but the result of a phase transition of what is actually, at very high temperatures, one single electroweak force. The standard model is able to describe the interactions of elementary particles in the framework of a field theory which is consistent with special relativity as well as quantum theory, emphasizing the important and elegant principle of invariance under local gauge transformations. In such a theory, interactions between particles are described by the exchange of bosonic field quanta. The calculation of experimentally observable quantities is usually done with perturbative methods, which can be used when

the relevant energy scale of the interactions is such that the coupling between the matter–particles and the particles that carry the forces is small. At energies typical for current accelerators, the experimental verification of the standard model, and in particular the electroweak sector, is excellent.

However, at the same time, the limitations of the standard model become more and more clear. There exist a large number of free parameters in the model, whose values are not predicted by the theory, but whose values are nevertheless very important for the world as we see it today, and for the consistency of the theory itself. It is not clear why there seem to be three families of quarks and leptons, nor why their masses are what they are. The evolution of the strength of the coupling of the force–particles to the matter–particles gives rise to the suspicion that in fact all forces in the current universe are the remnants of a single force at ultra-high temperatures, but the standard model does not include such a unification. A practical problem lies in the limited usefulness of perturbative methods when the forces are strong.

In the past decades, two methods have been extensively used by experimental physicists in order to undermine the standard model: high precision tests of unambiguous standard model predictions, and the search for exotic phenomena or particles that are not part of it. At the moment of writing of this thesis, neither have been able to measure any contradiction with the theory. In addition to these methods, it has always been a good experimental tradition to gather as much data as possible on phenomena that are not perturbatively calculable within the standard model, in the hope that they may contribute to a better understanding in the future.

Within the standard model, the theory of Quantum Chromodynamics (QCD) stands as the paradigm for the description of the strong nuclear interactions, just like Quantum Electrodynamics for electromagnetism. In QCD, the strong force has a symmetry that is mathematically described by the SU(3) group, with a triple charge named *colour*. The hadronic constituents of matter, the *quarks*, are colour triplets. They interact by exchanging the QCD field bosons called *gluons*, which are colour octets. The observation that the gluons carry colour results in the fact that they may interact with each other. The forces between quarks decrease when the distance between quarks gets smaller, at very small distances the quarks essentially do not feel each other. Vice versa, the forces get larger at increasing quark distance, up to the point that there is enough energy available to create new quark-antiquark pairs from the vacuum: the individual quarks can never be free, but 'dress up' as observable particles, the hadrons. A hadron, such as a proton, can thus be modeled to consist of a number of valence quarks, plus a sea of gluons, quarks and anti-quarks. QCD is the least well experimentally tested interaction in the standard model: due to the fact that QCD forces are strong, the coupling constant is relatively large at typical energies achieved so far, and perturbative methods fail, whereas a range of very high energies are needed for a precision test.

One of the areas that are not completely described by perturbative QCD is the structure of hadrons in terms of quarks and gluons. This structure is described by

a set of *structure functions*, which parametrize the probability and structure of the various interactions of a hadron. These structure functions can be interpreted in QCD. They depend, for a given hadron and a given interaction, on the resolution Q^2 with which the hadron is examined, and a variable x that can be interpreted as the fraction of momentum of the hadron carried by the quark or gluon that is being investigated. QCD is able to describe the evolution of these structure functions when the resolution Q^2 is varied. However, the actual shape of these functions at a given resolution is not determined by QCD, but needs to be measured by experiments.

Examples of processes that involve a very strong coupling between quarks and gluons are their fragmentation and hadronization, and the multiple radiation of soft gluons from quarks. Such processes have traditionally been described in a number of models. It is possible to speak of these models as 'QCD–inspired', since they do not actually follow from QCD, but contain elements and ideas that are compatible with it. Multiple gluon radiation with interference and quark–antiquark pair production from gluons, for example, is modeled in various QCD cascade models such as parton showers or colour dipole radiation.

In 1992, a new member of the accelerator family, the electron–proton collider HERA, has started delivering luminosity to two experiments, H1 and ZEUS. HERA is designed to perform precision tests of QCD and measure the structure of hadrons. It continues a long tradition of *deep inelastic scattering* (DIS) experiments, where the structure of hadrons is probed by leptons. Leptons are, in the standard model, pointlike and not subject to the strong interaction, and thus provide a clean way of penetrating nuclei and nucleons in order to measure their structure. It is convenient to think of such deep inelastic scattering facilities as a microscope: the object to be investigated is a proton, and it is being investigated by irradiating it using a beam of electrons. As in all microscopes, the smallest details that can be discerned are limited by the wavelength of the probe. In previous deep inelastic scattering experiments, electrons, muons or neutrinos were brought to scatter on a stationary target, and they have greatly contributed to what we know about QCD and hadrons today. However, a deeper insight into the hadronic structure can only be obtained with a better microscope, with smaller wavelengths. This wavelength is inversely proportional to the momentum transfer $\sqrt{Q^2}$ between lepton and hadron that determines the resolution with which the hadron is probed; HERA provides a very high maximum Q^2 by using, for the first time, high energy colliding electron and proton beams. By measuring structure functions at HERA, one will be able to test QCD by measuring the evolution of the structure functions with Q^2, *and* measure the hadronic structure in a completely new kinematical regime that has never been reached before. In addition, the measurement of the hadronic final state will provide new information on processes that are described by the mentioned QCD–inspired models. HERA will also open up a new kinematical region at low values of Bjorken x, where a breakdown of the standard Altarelli–Parisi evolution of structure functions is expected. Although values of x that can be reached at HERA

are probably still too high for a full investigation of these effects, HERA will offer a first and exciting glance.

Performing an experiment at HERA puts severe demands on the experimental techniques and the detector. In order to provide sufficient luminosity, HERA operates with 200 bunches per beam, spaced 96 ns apart. The resulting background rates of $\mathcal{O}(50)$ kHz, the fact that only a small fraction of this rate is due to DIS events, and the small time between two bunch crossings, imply that a very sophisticated data acquisition and trigger system is needed. In contrast to fixed target DIS experiments, not only the measurement of the scattered lepton, but also of the hadronic final state is important, for structure functions as well as for other QCD measurements. The ZEUS experiment at HERA has responded to these challenges by building an advanced detector, with a high resolution uranium/scintillator calorimeter for the measurement of the hadronic energy flow and for triggering as core.

In 1992, HERA has delivered only 1/1000th of its nominal (design) luminosity. Also the ZEUS detector was not complete. However, the calorimeter operated very well, and together with the fact that a modified trigger system was well able to select DIS events from background, this has already resulted in new and exciting data. The precision tests at high Q^2 will have to wait for more luminosity, but a new kinematical region at low x is already accessible, and also the test of fragmentation and multiple gluon radiation models can be performed.

In this thesis, we will the describe how the hadronic energy flow at HERA can be measured with the ZEUS calorimeter, and how this can be applied in an analysis of the hadronic final state, with emphasis on the testing of structure function parametrizations at low x, the testing of the QCD cascade models, and the observation of higher order processes that are a direct proof of the importance of gluons in deep inelastic scattering at HERA. This thesis has two focus points, which correspond to the main contributions of the author of this thesis to the ZEUS experiment: the interpretation of the raw calorimeter signals in a reconstruction program, and the analysis of the 1992 neutral current DIS event sample. Before doing so, we first describe the apparatus with which the measurements are performed.

The structure of this thesis is as follows. In chapter 2 we will review the formalism of deep inelastic scattering and structure functions, and the role of QCD in this, and we will give an overview of the physics that can be studied at HERA. We will discuss the signature of deep inelastic neutral current scattering events at HERA, and the methods of reconstructing from the measurements the relevant kinematical variables that characterize the ep collision. We will also show the predictions that exist for the neutral current DIS cross section, and the uncertainties that influence it.

Chapter 3 and 4 are devoted to the accelerator and detector with which the measurements are performed.

In chapter 3 we will shortly describe the HERA collider, how the luminosity is measured, and what backgrounds are induced by the beams.

The ZEUS detector, and in particular the high resolution calorimeter, will be described in chapter 4. We will pay attention to calorimetry, i.e. the measurement of particle energies by destructive methods. We will discuss how the knowledge of calorimetry has influenced the design of the ZEUS calorimeter, and we will describe the layout of this important detector component. We will explain how the calorimeter is read out, how the deposited energy is measured, and how it is calibrated.

Chapter 5, 6 and 7 are devoted to the software that has been written in the off-line environment in order to interprete the detector signals, reconstruct meaningful objects, and use them in a physics analysis.

In chapter 5, the off-line computing environment is described. After introducing the data management program ADAMO and its use in ZEUS, we shortly review a number of programs that have been written for detector and trigger simulation, and event reconstruction.

The event reconstruction software for the calorimeter interpretes the raw calorimeter data and reconstructs meaningful objects that can be used in a physics analysis. It has been designed, written and tested by the author of this thesis, and will be extensively described in chapter 6. We will pay attention to the reconstruction of the energy deposited in the individual cells. We will show how these cells can be clustered by a number of algorithms into objects that are intended to either correspond to showers of single particles, or to jets, the result of the fragmentation and hadronization of single partons. We will show how these objects are classified, and we will evaluate the efficiency of these procedures.

Chapter 7 is devoted to the simulation of *ep* interactions by Monte Carlo event generators. We will describe a number of models that are used for this simulation, and their implementation in event generator programs.

In chapter 8, 9 and 10 we describe the application of this software to an analysis of the hadronic final state in deep inelastic scattering events at HERA, measured with the ZEUS calorimeter. This analysis has been carried out by the author of this thesis as a parallel and independent second analysis to a similar analysis carried out at DESY, which has resulted in two published ZEUS papers [97, 102].

In chapter 8 we will describe in what conditions HERA and ZEUS have operated in their first year of luminosity (1992), we will describe the 1992 data sample, and the selection of deep inelastic neutral current events.

In chapter 9 we investigate the properties of the hadronic final state, and in particular the hadronic energy flow, in these selected events. We compare the data to the predictions of several Monte Carlo generators, and we will show which models best agree with the data. Also we will investigate the influence of the structure function parametrization.

In chapter 10, we will try to find jets in the data, and investigate the jet properties. In particular we will look for multi-jet events, as these are indications for the

occurrence of higher order processes in deep inelastic scattering.

Chapter 2

Deep Inelastic Scattering and HERA

Lepton-hadron scattering has proven to be a wealthy source of information on the structure of nuclei and nucleons for more than 25 years now [1, 2, 3, 4, 5]. Since leptons do not interact strongly, they are able to penetrate the nuclear surface and probe constituents with electric or weak charge. Illustrious experiments, like SLAC-MIT, EMC, NMC, BCDMS, CDHSW or CCFR, have greatly contributed to our knowledge of the structure of nucleons, and of the forces that keep them together. Traditionally, beams of charged (electrons or muons) or neutral leptons (neutrinos) have been brought to scatter on a stationary target. The virtuality Q^2 of the intermediate particle exchanged in the collision (a γ, a Z^0 or a W) determines the resolution with which a nucleus or a nucleon is probed. At high enough virtuality this evolves into a probe of the partons inside the nucleons.

2.1 Kinematics of DIS

We describe the kinematics of a general lepton-hadron interaction. The incoming lepton has 4-momentum k, and after the collision k'. The initial hadron, with 4-momentum p can keep its identity (elastic scattering) or break up into new particles. We denote the sum of the 4-momenta of the final hadronic state as p'. To lowest order in perturbation theory, we can describe the interaction to be mediated by exchange of an intermediate boson, as shown in figure 2.1.

The interaction is in first instance characterized by the centre-of-mass energy $\sqrt{s}$:

$$s = (k + p)^2 \tag{2.1}$$

where a standard four-vector quadrature with metric $(1, 1, 1, -1)$ is implied. s is thus a constant determined by the energy of the beams. In the evaluation of this

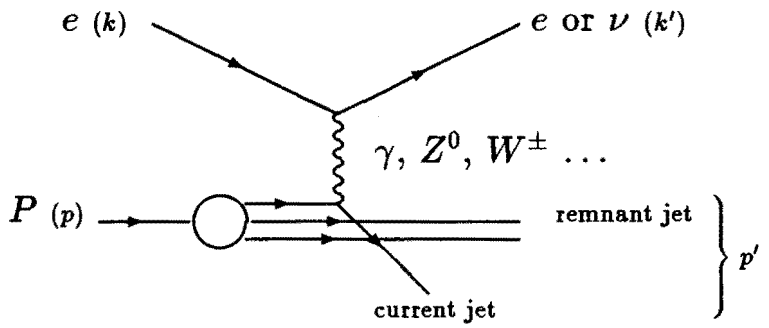

Figure 2.1: *Lowest order Feynman diagram of a lepton proton interaction. Mediators are a γ or Z^0 for neutral current interactions and a $W^{\pm}$ for charged current interactions.*

equation, beam particle masses can usually be neglected.

More interesting for a given interaction is the virtuality of the exchanged boson, i.e. the four-momentum transfer in the collision:

$$q^2 = (k - k')^2 \tag{2.2}$$

Since this is a space-like quantity, one usually prefers to take

$$Q^2 = -q^2 \tag{2.3}$$

which is always positive and runs between 0 and s. Q^2 sets the scale of the interaction, as it describes what the exchanged boson sees of the proton. At high Q^2, the lepton-proton interaction is essentially a lepton-quark interaction. In lowest order, this leads to a *current jet* from the struck quark, and a proton (target) remnant jet.

The Lorentz invariant quantity

$$y = \frac{p \cdot q}{p \cdot k} \tag{2.4}$$

runs between 0 and 1. In the proton rest frame this quantity reduces to the fractional energy transfer from proton to lepton.

The Bjorken scaling variable x (*elasticity*) is defined as:

$$x = \frac{Q^2}{2p \cdot q} \tag{2.5}$$

and plays an important role in deep inelastic scattering and the theory of structure functions. The region of deep inelastic scattering is characterized by large values of Q^2: $Q^2 \gg 1$ GeV^2. In the parton model, which will be described later, x can be

interpreted as the fraction of the proton momentum (formally, in a frame in which the proton has infinite momentum) carried by the struck quark. It is thus easily seen that x runs between 0 and 1.

The variables x, y and Q^2 are not independent, but related through

$$Q^2 = xys \tag{2.6}$$

Any two of the set are sufficient to describe the interaction, excluding the details of the hadronic final state.

The total invariant mass of the final hadronic system is given by:

$$W^2 = (p')^2 = (p+q)^2 = Q^2\frac{1-x}{x} + m_p^2 = sy - Q^2 + m_p^2 \tag{2.7}$$

where m_p^2 can usually be neglected.

When the interaction is purely electromagnetic, the exchanged boson is a photon. At higher values of Q^2 ($Q^2 > 1000$ GeV^2), the electroweak theory predicts an increasing importance of the weak interaction and the interference between electromagnetic and weak interactions. In neutral current interactions, whether mediated by photons or Z^0, the scattered lepton retains its identity. For charged current interactions, the exchanged particle is a W, and charged leptons are transformed into neutrino's and vice versa.

2.2 Cross Sections

To lowest order, a lepton-hadron collision can be depicted as shown in figure 2.1 (with the hadron specified to be a proton, and the lepton an electron), and the cross section for this interaction can in general be written as [1]:

$$d\sigma = (flux\,factor)L_{\mu\nu}W^{\mu\nu}d(phase\,space) \tag{2.8}$$

with $L_{\mu\nu}$ the lepton tensor, describing the leptonic side, and $W^{\mu\nu}$ the hadron tensor, describing the hadronic side.

In the standard model, leptons are pointlike. This means that the contribution to the cross section of the lepton vertex is relatively simple and straightforward.

The hadron tensor, however, is complicated and contains all information on the structure of the hadron current. It is well known that hadrons are not pointlike. In the evaluation of the cross section, this is reflected by the introduction of *structure functions*, which implicitly describe the hadronic structure. They depend on two independent variables, for which we usually take Bjorken x, and Q^2. In deep inelastic scattering, the number of relevant structure functions is limited to 3.

We now turn explicitly to the case of electron-proton scattering. We assume the possibility of left and right handed electrons and positrons as incoming leptons.

We can now write the total inelastic cross section for neutral and charged current scattering in bins of x and Q^2 in terms of the 3 structure functions F_i:

$$\frac{d^2\sigma(e^{\mp}_{L,R}p)}{dx\,dQ^2} = \frac{4\pi\alpha^2\mathcal{P}^2(Q^2)}{x}\Big[y^2 x F_1^{L,R}(x,Q^2) + (1-y)F_2^{L,R}(x,Q^2) \pm (y-\frac{y^2}{2})xF_3^{L,R}(x,Q^2)\Big] \quad (2.9)$$

with L and R denoting the handedness of the incoming electrons, and α the electromagnetic coupling constant. The propagator factor $\mathcal{P}^2(Q^2)$ depends on the type of boson exchanged in the interaction, and has therefore been taken out explicitly. It takes the following form:

$$\mathcal{P}^2(Q^2) = \begin{cases} \frac{1}{Q^4} & \text{single } \gamma \text{ exchange,} \\ \frac{1}{(Q^2+M_Z^2)^2} & \text{single } Z^0 \text{ exchange,} \\ \frac{1}{Q^2(Q^2+M_Z^2)} & \gamma\text{–}Z^0 \text{ interference term,} \\ \frac{1}{(Q^2+M_W^2)^2} & \text{single } W^{\pm} \text{ exchange.} \end{cases} \quad (2.10)$$

From the expressions for the propagator it can be seen that neutral current scattering at low Q^2 is mostly electromagnetic, with the contributions from γ–Z^0 interference and Z^0 exchange suppressed by M_Z^2 and M_Z^4 respectively. Similarly, the cross section for charged current scattering peaks at high Q^2.

It should be noted that the structure functions F_i^{weak} for the weak interactions, as probed in neutrino scattering, differ from the structure functions F_i^{em} for electromagnetic interactions, as probed in electron or muon scattering, and are different for charged and neutral current interactions. A photon couples to the charge of the constituents, whereas a Z does not, and a W, exchanged in charged current interactions, couples to certain quark flavours only. We will elaborate on this in the next sections. As an example, structure function F_3 describes the parity violating term of the cross section, and therefore only takes part in weak interactions. F_3 equals zero for pure γ exchange.

2.3 The Quark-Parton Model

The information obtained in the first deep inelastic electron scattering experiments at SLAC have played a decisive role in the development and acceptation of the quark-parton model [6]. Already before these first experiments were done, current algebra methods had given rise to the expectation that the structure functions in deep inelastic scattering at high Q^2 would only depend on x, and not on Q^2 [7]. This phenomenon was called *scaling*, and was experimentally verified at SLAC [6]. It was soon realized that this scaling phenomenon implied substructure in the nucleon; the sub-nucleonic particles were named *partons* in general [8]. In the simple quark parton model, these partons are identified with quarks. In this model, deep inelastic

electron-proton scattering can be seen as elastic scattering of the electron on a point-like parton, with the rest of the proton as spectator. This interaction is insensitive to the photon virtuality Q^2, thus explaining scaling. The variable x can be interpreted as the momentum fraction of the proton carried by the struck quark. The final state results from the fragmentation of the partons involved in the interaction, i.e. the struck quark (or current quark) and the spectator quarks.

For spin 1/2 partons (like quarks) with no transverse momentum, the Callan-Gross equation [9] holds:

$$F_2(x, Q^2) = 2xF_1(x, Q^2) \tag{2.11}$$

which implies that the partons couple only to transversely polarized photons.

The partons q ($q = u, d, s, c$) are characterized by a momentum density function $q(x)$ in the hadron. Evaluating the ep cross section in this picture, one can write out the structure functions F in terms of these parton momentum density functions $q_i(x)$. We do this for neutral current scattering:

$$F_2(x, Q^2) = \sum_i A_i(Q^2)[xq_i(x) + x\bar{q}_i(x)] \tag{2.12}$$

$$F_3(x, Q^2) = \sum_i B_i(Q^2)[q_i(x) - \bar{q}_i(x)] \tag{2.13}$$

where the sum runs over all quark flavours i in the proton. The functions on the left hand side of the equations have contributions from pure γ exchange, γ/Z^0 interference, and pure Z exchange, which in the equations above are hidden in the coefficients A and B. For pure γ exchange $A_i(Q^2) = e_i^2$, i.e. the charge of the partons squared, and $B_i(Q^2) = 0$ since the electromagnetic interaction conserves parity [1]. A_i and B_i get extra terms that include the vector and axial vector coupling of the Z to fermions when pure Z exchange and the $\gamma - Z$ interference term play a role in the interaction (at high Q^2, as explained). In order to give more explicitly the contributions to a structure function, for example F_2, it is sometimes written:

$$F_2(x, Q^2) = F_2^{em}(x, Q^2) + F_2^{\gamma/Z}(x, Q^2) + F_2^Z(x, Q^2) \tag{2.14}$$

where the implicit contributions in A are written out explicitly. From the equations above, it is clear that for pure γ exchange, F_2 does not depend on Q^2.

2.4 QCD and Deep Inelastic Scattering

The quark-parton model involves a static picture of the proton. A full theory, however, should include the use of QCD as theory of the strong interactions for a

[1] These equations hold for charged lepton - proton scattering. The corresponding equations for charged lepton - neutron scattering are easily obtained assuming isospin symmetry, so that the u (d) distribution in the proton is the d (u) distribution in the neutron, while the s and c distributions remain unchanged. For neutrino-hadron scattering, the charge of the quarks do not play a role, and a W couples to certain flavours only, see [1, 4].

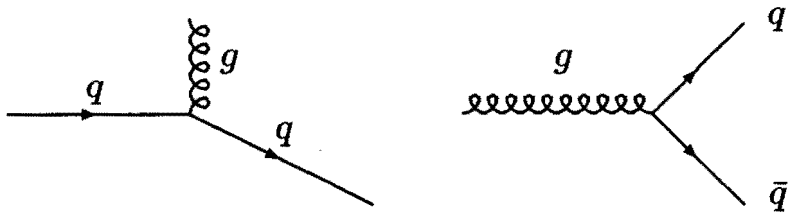

Figure 2.2: *Gluon bremsstrahlung by a quark (left) and quark-antiquark pair production by a gluon (right). Not shown is the vector-boson probing the resulting quark or antiquark.*

description of the dynamic structure of the proton. This implies a view in which the proton consists of bound, but asymptotically free, quarks, interacting through gluon exchange. Perturbative QCD is indeed able to describe and predict certain features of the structure functions. Two simple diagrams, as shown in figure 2.2, govern the leading order QCD corrections to the naive quark parton model: gluon bremsstrahlung by a quark, and quark-antiquark pair production by a gluon. The strength of the coupling between quarks and gluons is given by the strong coupling constants α_s. At low Q^2, the vector boson exchanged in the deep inelastic scattering is unable to resolve the individual partons, but at higher Q^2, the individual quarks and antiquarks that result after the gluon radiation or pair production can be seen. The gluon self-interactions, although also important, are not shown since the gluon does not interact with the vector boson.

Via these processes, the strength of the quark–gluon coupling α_s itself depends on the scale Q^2. Perturbative QCD is able to calculate the diagrams of figure 2.2, and therefore to predict this dependence: α_s will obtain a typically logarithmic Q^2 dependence, and is often referred to as a *running* coupling constant. In leading order:

$$\alpha_s(Q^2) = \frac{12\pi}{(11N_c - 2N_f)\ln(\frac{Q^2}{\Lambda^2_{QCD}})} \tag{2.15}$$

where $N_c = 3$ is the number of colour charges, $N_f \leq 6$ is the number of open flavours in QCD, and Λ_{QCD} denotes a fundamental QCD parameter.

The interactions between quarks and gluons, and the interactions between the gluons themselves, also introduce a dependence of the quark distribution functions, and therefore of the structure functions, on the scale Q^2: the scaling is violated. Again, perturbative QCD is able to predict that this dependence will typically behave like $\ln Q^2$. It is convenient to divide structure functions into two types:

- The leading order evolution of *nonsinglet* structure functions F^{NS} is com-

pletely determined by the left diagram of figure 2.2. This makes the evolution equation for $dF^{NS}/d\ln Q^2$ relatively simple, since it only depends on the same structure function F^{NS}:

$$\frac{dF^{NS}(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}\int_x^1 F^{NS}(z,Q^2)P_{qq}(\frac{x}{z})dz \qquad (2.16)$$

where the integration is performed over all momenta z between x and 1, and where P_{qq} is a 'splitting function' that describes the left diagram of figure 2.2. In terms of quark density functions, nonsinglet structure functions are typically of the form $q_i - q_j$ or $q - \bar{q}$, such as F_3.

- The evolution for *singlet* structure functions is governed by both diagrams of figure 2.2, and is therefore more complicated. The gluon density is also important, and is subject to evolution with Q^2 itself. Singlet structure functions, like F_2, are typically of the form $q + \bar{q}$.

The evolution equations for the structure functions in deep inelastic scattering are known as the Altarelli-Parisi equations [10], and involve, as shown above for the nonsinglet case, besides the structure functions at a given scale also α_s, and calculable gluon and quark splitting functions that effectively describe quark–gluon and gluon–gluon diagrams.

Since perturbative QCD predicts the evolution of structure functions with Q^2, the measurement of this evolution is an important test of QCD. However, it needs a long lever arm in Q^2 due to the fact that the evolution goes only like $\ln Q^2$. A global QCD fit to the data can give α_s (or Λ_{QCD}), and, since the gluons play an important role in the quark density evolution with Q^2, the gluon distribution $G(x,Q^2)$. It should be noted, however, that the x-dependence of the structure functions is not predicted by QCD.

In the QCD picture, the Callan-Gross relation is slightly violated. The parameter R measures the amount of this violation, by giving the cross section for longitudinally polarized photons divided by the cross section for transversely polarized photons:

$$R = \frac{\sigma_L}{\sigma_T} = \frac{F_2 - 2xF_1}{2xF_1} = \frac{F_L}{2xF_1} \qquad (2.17)$$

with F_L the longitudinal structure function:

$$F_L(x,Q^2) = F_2(x,Q^2) - 2xF_1(x,Q^2) \qquad (2.18)$$

The transverse momentum of quarks, caused by the fact that they are subject to strong forces that bind them in the proton, gives rise to a non-zero value of R. At low Q^2, this is the dominating cause of a non-zero R. With the presence of non-spin-1/2 particles like gluons in the proton, also QCD predicts that diagrams like those shown in figure 2.2 will contribute to a non-zero R. Although they are not very large, at fixed x these contributions only fall slowly with Q^2. Therefore, measuring R (i.e. F_L) also gives information on the gluon content of the proton.

The methods of perturbative QCD as applied above in the description of the evolution of structure functions, are only valid at collision energies where bound kinetic and potential energies of the struck quark may be neglected. Such *higher twist* terms contribute to the evolution as powers of $1/Q^2$, and disturb the $\ln Q^2$ behaviour at low Q^2, but are negligible at high Q^2.

QCD not only has effects in the momentum distributions of struck partons, but also in the processes of hadronization and jet forming, as eventually measured in the detector. We will come back to these effects later.

2.5 Existing DIS Data

Data on structure functions and hadronic final states have been obtained in several fixed target DIS experiments at SLAC, FNAL and CERN, with electron, muon, and neutrino beams [2, 3, 4]. The kinematical range explored by these experiments is limited by the beam energies and the requirement that the final state must be measurable, which means $10^{-2} < x < 0.8$ and $0.1 < Q^2 < 200$ $(\mathrm{GeV}/c)^2$. Apart from proton and neutron structure functions, *nucleon structure functions* are often used, which average the proton and neutron structure functions. They are in fact the measured structure functions on heavy isoscalar (equal number of protons and neutrons) targets. Using data from neutrino and antineutrino scattering on hydrogen or deuterium, the individual quark density functions can be determined. Experiments with electrons or muons need hydrogen as well as deuterium targets, and additional assumptions on isospin symmetry and antiquark densities, but have much better statistics. Figure 2.3 summarizes the data on the proton structure functions F_2 from SLAC (e) and BCDMS (μ) [11]. Apart from structure functions, results on nuclear effects, sum rules, and QCD analyses have been published.

The results of the experiments have been used in parametrizations of the structure functions that have been made available in program libraries [12, 13]. Some of the more recent parametrizations include those of Morfin and Tung (MTB1, MTB2) [14], and Martin, Roberts and Stirling (MRSD0, MRSD-) [15].

2.6 Physics at HERA

We take units such that $c = \hbar = 1$. The recently commissioned HERA electron-proton collider will take further important steps in deep inelastic scattering. Details of the HERA machine will be described in the next chapter. We will limit ourselves here to the information that electrons are accelerated to 26.7 GeV and made to interact with 820 GeV protons coming from the opposite direction. The centre-of-mass energy in this configuration is:

$$\sqrt{s} = 2\sqrt{E_e E_p} = 296\,\mathrm{GeV} \tag{2.19}$$

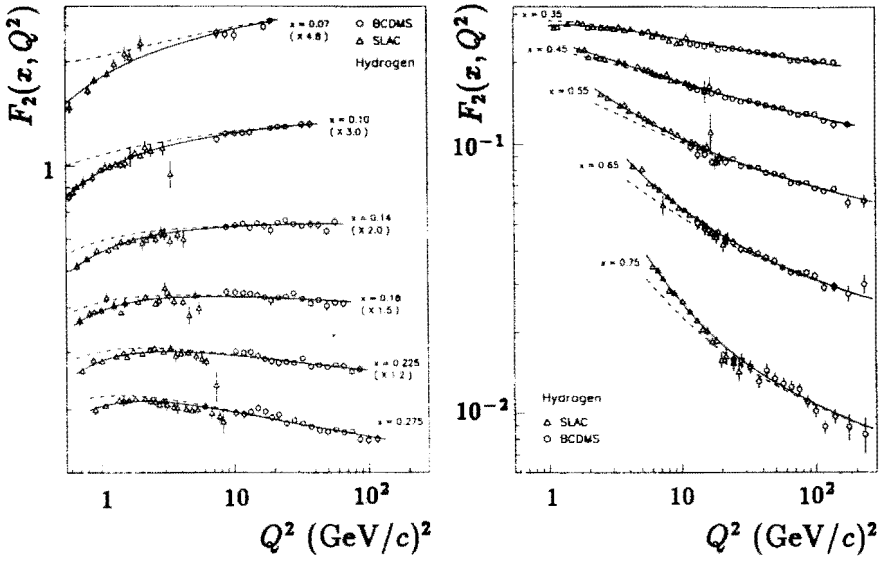

Figure 2.3: *The structure function F_2 of the proton as measured in fixed target electron and muon scattering. The lines represent a next-to-leading order QCD fit with (solid line) and without (dashed line) corrections for higher twist effects.*

and the maximum Q^2 in interactions at HERA is:

$$Q^2_{max} = s = 87600\,\mathrm{GeV}^2 \tag{2.20}$$

more than 2 orders of magnitude larger than what has been obtained before. At such high values of Q^2, ep-interactions are essentially electron–quark interactions. HERA will also extend the x-range of measurements in deep inelastic scattering down to $x \approx 10^{-4}$, again 2 orders of magnitude smaller than what has previously been obtained.

The large asymmetry in the beam momenta has consequences for the kinematics of the final state. These, and their experimental aspects, will be discussed in detail in section 2.7.

We will give a brief overview of physics opportunities at HERA. An excellent extended description of these opportunities can be found in reference [16].

2.6.1 Electroproduction

The deep inelastic region at HERA can be defined as the kinematic region where the scattered electron enters in the main detector. For neutral current interactions, such processes are sometimes also denoted as *electroproduction*. The scattered electron enters in the main detector if Q^2 is above a few GeV^2. We take $Q^2 = 4\,\mathrm{GeV}^2$ as lower cut-off. This value coincides with the lower end of the region of validity of many

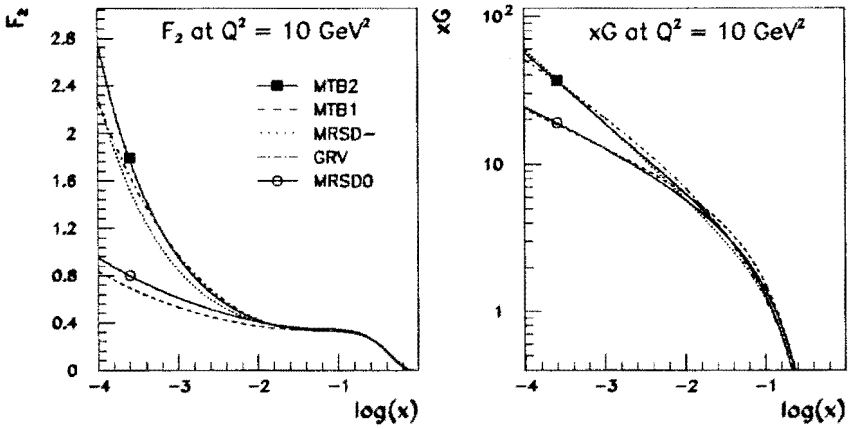

Figure 2.4: *Behaviour of F_2 and the gluon distribution function G as a function of x, at Q^2 = 10 GeV^2, as predicted by some recent parametrizations. The legenda in the left plot also holds for the right plot.*

structure function parametrizations. The directions and energy of scattered lepton and quark as a function of x and Q^2 are more closely investigated in section 2.7.

Integrating equation 2.9 over kinematical and helicity variables, and neglecting R, we obtain the total cross section (without radiative corrections) for deep inelastic scattering at HERA, as predicted by the standard model. This calculation, however, needs the previously mentioned parametrizations of structure functions as input, since these are not predicted by QCD (QCD only predicts their evolution in Q^2). These parametrizations are based on existing data, but a large part of the kinematically allowed region at HERA, notably the low x region, is not covered by this data. The parametrizations therefore need to be extrapolated, with considerable theoretical uncertainties, into this new region. This makes the predictive power and the reliability of these parametrizations in this new region poor, and several parametrizations may differ considerably. Some of these differences in F_2 and the gluon distribution $G(x, Q^2)$ are illustrated in figure 2.4. The differences in F_2 have an immediate influence in the total cross section for deep inelastic scattering.

We have calculated the total cross section for neutral current and charged current scattering at HERA in various bins of x and Q^2, for two different parametrizations, using the Monte Carlo program LEPTO [17]. The results are shown in figures 2.5 and 2.6. The two parametrizations are Morfin-Tung B1, with a fairly flat gluon distribution at low x, and Morfin-Tung B2, with a steeply rising gluon distribution at low x [14]. Since these cross sections have been calculated with a Monte Carlo method, each bin has a statistical error of approximately 1%. It should be noted that radiative corrections (see section 2.6.3) have a considerable influence in certain bins.

The neutral current cross section peaks at low Q^2, whereas the charged current

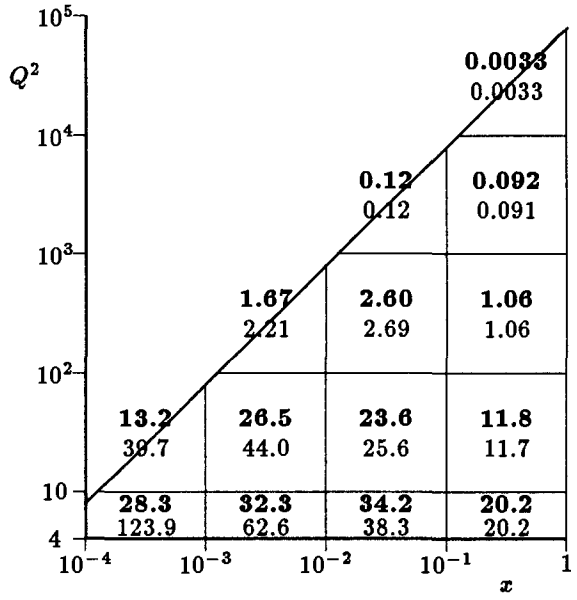

Figure 2.5: *Predicted cross section in nb for NC scattering at HERA (LEPTO, no radiative corrections), in shown bins of x and Q^2. Results obtained with the Morfin-Tung B1 structure function parametrisation are written fat above the results obtained with MTB2. For a discussion, see text.*

cross section reaches its highest values at higher Q^2, due to the M_W^2 term in the propagator. The cross sections obtained with MTB1 and MTB2 differ considerably at low x, low Q^2, as expected. HERA experiments will be able to distinguish between the two parametrizations, although detector effects will smear out the large differences to some extent. Without radiative corrections, the total neutral current cross section for $Q^2 > 4$ GeV^2 equals 195 nb for MTB1 (372 nb for MTB2), for $Q^2 > 10$ GeV^2 this is 80 (127) nb, and for $Q^2 > 100$ GeV^2 this number equals 5.5 (6.1) nb. For charged current scattering, these numbers are 69.6 (71.9) pb, 69.3 (71.3) pb, and 64.9 (65.2) pb respectively.

As R is small in most of phase space covered by HERA, the contribution to the cross section of F_1 is small. The contribution of F_3 comes up at high Q^2, first via the γ–Z interference term, then via pure Z exchange. Its measurement will need a considerable amount of data, whereas F_2 can be measured to a large extent with less data. A rough idea of what can be done with 100 pb^{-1} in measuring F_2 is given in figure 2.7. HERA can provide a measurement between $Q^2 \approx 10$ GeV^2 and

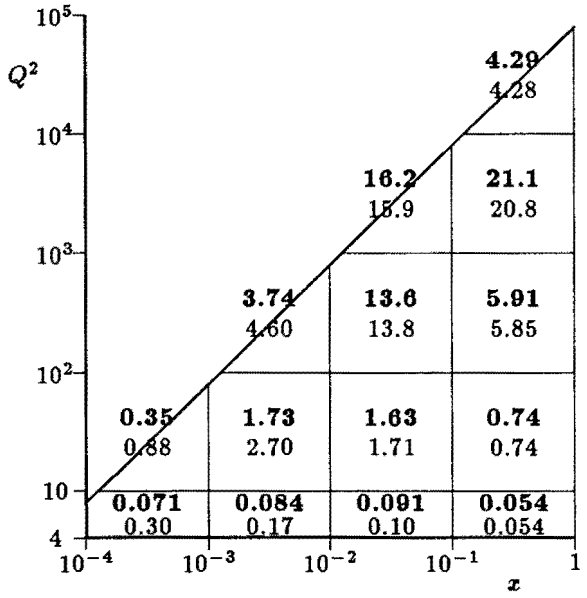

Figure 2.6: *Predicted cross section in pb for CC scattering at HERA (LEPTO, no radiative corrections), in shown bins of x and Q^2. Results obtained with the Morfin-Tung B1 structure function parametrization are written fat above the results obtained with MTB2. For a discussion, see text.*

$Q^2 = 40000$ GeV^2, and between $x \approx 10^{-4}$ and $x = 0.7$.

With the nominal beam energies, HERA data will have little overlap with existing DIS data. In order to solve this, it has been proposed to take additional data at a lower centre-of-mass energy (e.g. $s = 12000$ GeV^2), as also shown in figure 2.7. This would considerably increase the overlap, which is useful for reduction of systematic errors.

In order to unfold the individual quark distribution functions, HERA has several options in changing the beams. When positrons are accelerated instead of electrons, the total unpolarized neutral current cross section differs from the electron–proton cross section when γ/Z interference and Z exchange play a role. The charged current cross section differs even more, since different quark flavour distributions are probed in positron scattering compared to electron scattering. In addition, the lepton beam can be longitudinally polarized (see section 3.1). This implies the selection of left handed or right handed electrons or positrons. The coefficients A and B in equations 2.12 and 2.13 are sensitive (through the weak interactions) to

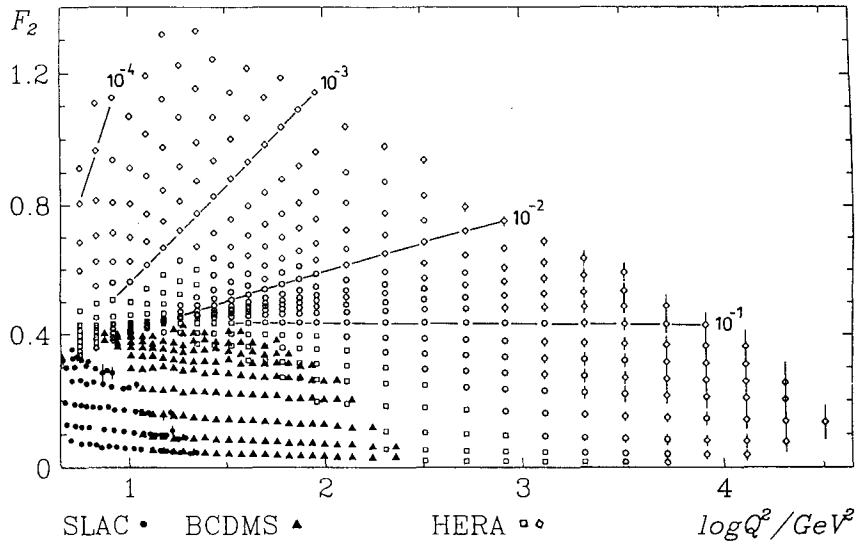

Figure 2.7: *F_2 measurement at HERA with 100 pb^{-1} at the nominal value of s, and 100 pb^{-1} at a lower s of 12000 GeV^2. Some existing data on F_2 is shown at the left.*

this polarization. For charged current scattering, $\sigma_{e_L^-} > \sigma_{e_R^+}$ and the peculiar effect occurs that $\sigma_{e_R^-} = \sigma_{e_L^+} = 0$ due to the coupling of the exchanged $W^\pm$. Furthermore, deuterons can be accelerated instead of protons, which will facilitate the measurement of the quark density combinations of the neutron. A global QCD analysis of the deep inelastic scattering data can give information on Λ_{QCD}.

In an analysis of HERA deep inelastic scattering data, higher order processes (compared to the first order process shown in figure 2.1) need to be taken into account. Two of such higher order processes are shown in a Feynman diagram in figure 2.8. The process in the left diagram is known as boson-gluon fusion (in this case photon-gluon fusion). The exchanged boson interacts with a gluon in the proton via the production of a quark-antiquark pair. These will give rise to two back-to-back (in azimuth ϕ) jets in the final state. Although of higher order, this process dominates the activity in the detector at low Q^2 over the first order process via the produced quark-antiquark pair. The scale is not set by the virtuality Q^2 of the exchanged boson, but is closer related to the invariant mass of the $q\bar{q}$ system. The process in the right diagram, the initial or final state radiation of a hard gluon, is usually called QCD-Compton scattering in the case of $Q^2 \sim 0$ (a real or almost real photon).

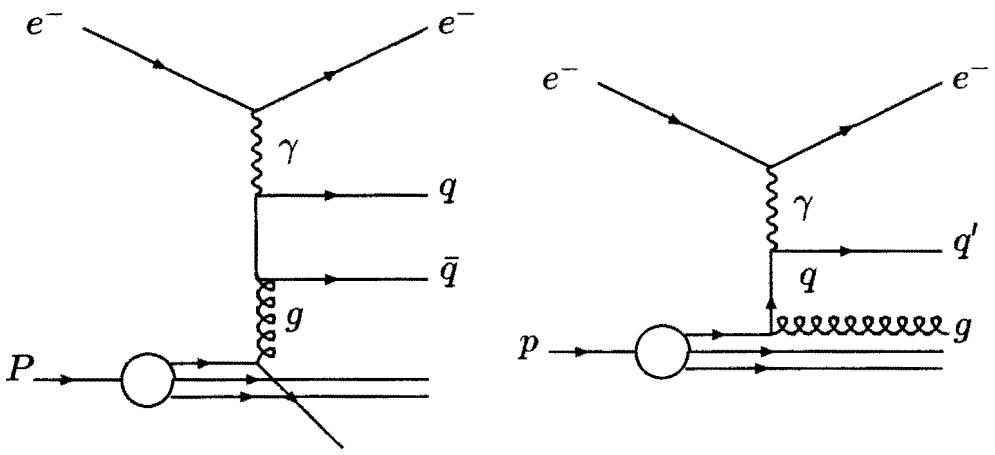

Figure 2.8: *First order corrections to the standard deep inelastic scattering process include boson gluon fusion processes and gluon radiation. The crossed diagrams for BGF and final state gluon radiation need to be taken into account as well.*

2.6.2 Hadronic final states and jets

Quarks and gluons taking part or being produced in the hard parton scattering hadronize into particles that can be observed in the detector. One can distinguish several phases in the hadronization process, corresponding to different time scales, as shown in figure 2.9:

1. The hard scattering of partons (photons, quarks, gluons), which can be described by QCD and electroweak theory;
2. 'Hard' gluon radiation leading to distinct jets, which can in principle be described perturbatively in QCD, and which happens on a typical distance scale of 10^{-15} cm;
3. Small angle soft gluon radiation and gluon splitting ($\sim 10^{-14}$ cm), which happens on a low virtuality scale, where α_s becomes large and perturbative methods start to loose their validity;
4. Quarks and gluons combine to hadrons ($\sim 10^{-13}$ cm), which is in principle governed by QCD, but can not be calculated perturbatively;
5. Unstable hadrons decay into experimentally observable particles.

At HERA, the final state is sensitive to both initial and final state QCD radiation effects, from hard as well as from soft gluons. This is different from the situation at

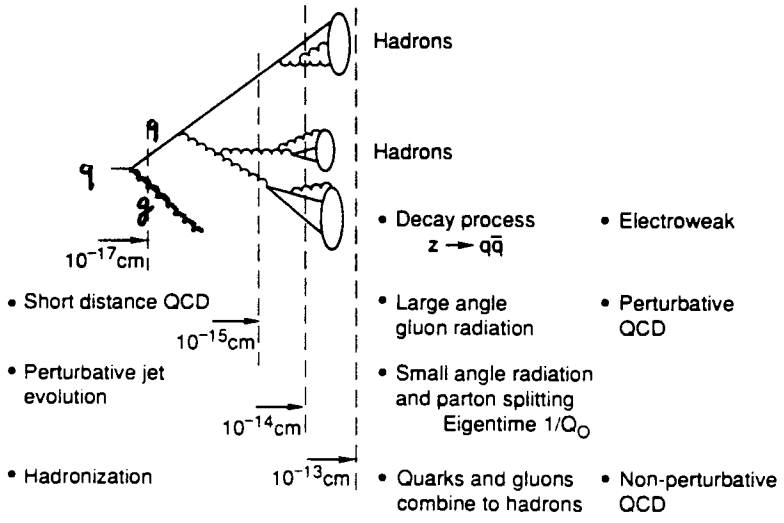

Figure 2.9: *The stages and time scales in the evolution of a jet.*

LEP, where there are no incoming strongly interacting particles, but the situation at HERA is much cleaner than the final state in $p\bar{p}$ interactions. There is a considerable interest in the mentioned QCD effects, which include hard parton emission giving rise to jets, semihard perturbative effects such as multiple gluon emission with colour coherence effects, and non-perturbative phenomena like hadronization and confinement. Since one balances at the edge of validity of perturbative methods, exact matrix element calculation and QCD-inspired models merge, which makes this an exciting topic. The various models have been implemented in a number of different Monte Carlo programs that will be described in chapter 7. A comparison between these models and HERA data in a completely new kinematical regime will be presented in chapter 9.

The hadronic final state has not been very well studied in fixed target DIS experiments, due to the experimental difficulties involved. HERA will be the first deep inelastic physics facility where the energy is large enough to allow clear jet signatures in the final state. We will explicitly describe the jet reconstruction methods in section 6.9. The jets at HERA are interesting in themselves and provide direct ways of accessing the hard scattering processes. Single jets at high transverse energy reflect the properties of struck partons. The energy flow within a jet is governed by the soft processes of small angle gluon radiation and fragmentation. Multi-jet production in DIS events could be evidence for the importance of higher order processes as shown in figure 2.8, which have not been studied before, and which are

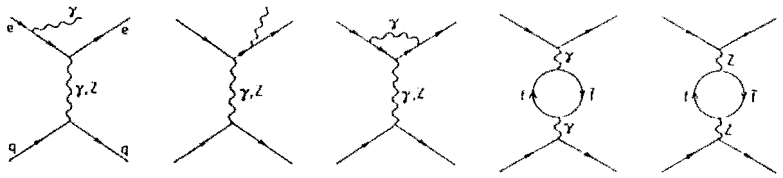

Figure 2.10: *The $\mathcal{O}(\alpha)$ leptonic corrections for neutral current scattering and the one-loop fermionic contributions to the photon and Z self energies*

directly related to the role of gluons in DIS at HERA. We will elaborate on this in chapter 10.

From the relative amount of 2-jet and 3-jet events in the data, α_s can be determined. By doing this at various values of Q^2 (a unique feature of HERA), the running of α_s can be proven within a single experiment.

2.6.3 Radiative Corrections

Higher order electroweak effects change the amplitudes of the processes described above. Moreover, they include processes that give rise to additional photons in the final state. The full radiative corrections include radiation from the lepton line as well as from the quark line and interference between leptonic and quarkonic radiation, and loop corrections that contribute to the self energy of the exchanged boson. It has been shown that the corrections are dominated by contributions describing the radiation of real and virtual photons from the lepton line [18]. These leptonic corrections, together with the one-loop fermionic contributions to the γ and Z self energies, as shown in figure 2.10 [19], are sufficient to describe NC scattering with an accuracy of better than 5 %, and even better than 1 % in a large part of phase space. The radiation of photons from the lepton line include initial and final state radiation, as well as vertex corrections. It has been shown that the radiative corrections to the cross section are large and negative at low y and high x, and large and positive at high y and low x [18, 20]. The corrections at very low x can be as large as 30%, and since the cross section peaks at low x, they have a similar influence on the total cross section.

2.6.4 Phenomena at low x

The deep inelastic regime is characterized by high Q^2. A relatively high Q^2 is needed to keep α_s small and thus retain the validity of the parton picture in perturbative QCD. QCD is able to describe the evolution of the structure functions with Q^2, but not their actual shapes at a given Q^2. Once these structure functions are measured, however, the QCD evolution equations predict their shapes at any x or Q^2, as long as the evolution equations are still valid at that particular kinematics. The

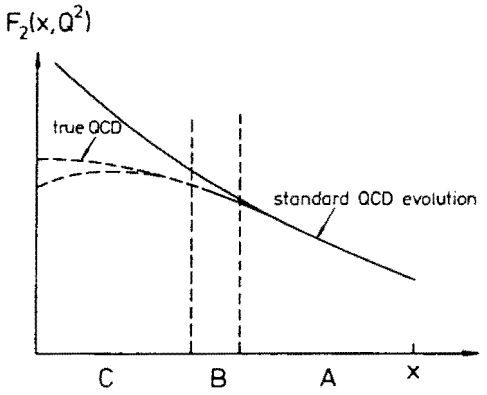

Figure 2.11: *Small-x behaviour of F_2. For an explanation, see text.*

QCD evolution in the deep inelastic region is well described with the Altarelli-Parisi equations.

At low x and low Q^2, on the other hand, we enter the non-perturbative region. At HERA, we have the unique possibility to go down in x to $x \approx 10^{-4}$, yet keeping Q^2 high: our parton model is still valid. The standard QCD evolution here predicts a steep rise of the structure functions that will eventually violate unitarity. It is expected that in this region additional parton interactions will occur that will eventually keep the structure functions finite by *screening* the partons from each other. Gribov, Levin and Ryskin have, based on a QCD inspired model, added a nonlinear term to the evolution equations. Figure 2.11 shows the predictions of this model on the behaviour of F_2 versus x at fixed Q^2, as compared to the standard evolution equations. A denotes the deep inelastic region where the standard perturbative evolution is valid. C is the non-perturbative region. In between lies B, where the parton picture still holds, but additional terms modify the QCD evolution. HERA may be able to enter the B region, in particular if screening effects already turn on locally in the proton (*hot spots*). The low x experiments focus on the inclusive electron energy spectrum, the measurement of structure functions such as the gluon distribution at low x, and the measurement of specific hot spot jet topologies and hadronic final states.

2.6.5 Photoproduction

At low values of Q^2, the scattered lepton disappears into the beam pipe, and cannot be measured in the main detector [2]. In these processes, the exchanged photon is real or almost (quasi) real, and we speak about *photoproduction*. From the propagator

[2]Part of these scattered electrons in neutral current events, however, can be measured in the electron calorimeter of the luminosity monitor, as will be explained in section 3.2.

term in the cross section equation, it can be seen that the cross section at HERA peaks for $Q^2 \sim 0$, and photoproduction events form the majority of HERA data, even up to such extent that they become background for more interesting physics. There is, however, considerable interest in photoproduction. The small Q^2 of the exchanged boson is not necessarily the scale of the interaction with the proton constituents. A typical example is the boson-gluon fusion process shown left in figure 2.8, where the scale is rather related to the invariant mass of the produced $q\bar{q}$ pair. In this way, hard interactions can occur with much larger transverse energy E_T than expected from a lowest order current jet that balances the scattered lepton. So far we have only dealt with direct photon interactions, but at $Q^2 \to 0$, the interaction is also sensitive to the internal structure of the photon. The photon can act like a hadron (vector meson) and have a hadronic interaction, usually with low p_T, but with a high cross section. In addition a parton (q, $\bar{q}$, g) inside the photon can interact with a parton inside the proton, leaving a photon remnant. Such processes are often referred to as *resolved* photon interactions. The total γP cross section has been measured by ZEUS to be $154 \pm 16 \pm 32$ μb [21], with the majority of soft (vector meson dominance-like) and resolved photon interactions. The total cross section in lowest order for boson-gluon fusion with $M_{q\bar{q}} > 3$ GeV equals approximately 2 μb. Again, accurate jet reconstruction can give a direct handle on the hard scattering processes. Eventually, the parton momentum distributions in the photon might be obtained from resolved photon processes.

2.6.6 Heavy Quark Production

Heavy quarks can be produced in electroproduction as well as in photoproduction, but clearly the cross section dominates at low Q^2. Boson-gluon fusion is a copious source of heavy quarks: with $M_{q\bar{q}} > 3$ GeV, the lowest order cross section for $c\bar{c}$ production equals 430 nb, whereas the $b\bar{b}$ cross section equals 4.3 nb. Also resolved photon processes contribute to heavy quark production. At HERA, one can study rare c and b decays, and extract the gluon distribution in the proton via the detection of J/Ψ and open charm production.

2.6.7 Electroweak physics

Testing of the electroweak part of the standard model is not one of the primary aims of HERA. Nevertheless, at HERA one can study $R = \sigma_{CC}/\sigma_{NC}$, which is a measure of M_W, the $WW\gamma$ vertex, and one can search for new vector bosons and right handed currents. For example: Z' masses in the range from 400 GeV to 800 GeV can be probed, and heavy Majorana neutrinos with masses up to 160 GeV can be found at HERA [22].

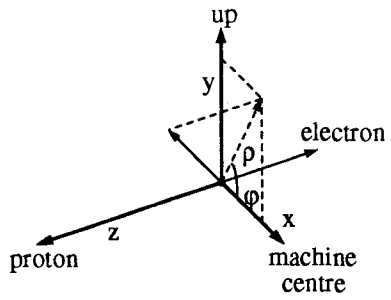

Figure 2.12: *Global HERA and ZEUS coordinate system.*

2.6.8 Beyond the Standard Model

HERA is well suited to search for particles that can be created in the s-channel in ep collisions. Such particles would carry lepton number as well as colour, and are known as *leptoquarks* or *leptogluons*. At HERA they can be found up to a mass of 300 GeV; based on an analysis of 1992 data, ZEUS has ruled out, at the 95% confidence level, scalar isosinglet leptoquarks with electroweak coupling strength to $(e^- u)$ states for masses below 168 GeV for left-handed couplings, and below 176 GeV for right-handed couplings [23]. Also supersymmetric particles (especially if R-parity is not conserved) and excited/composite leptons and quarks can be detected.

2.7 Reconstruction of the Kinematics

In the remainder of this thesis, we will describe HERA and ZEUS in a global coordinate system, as shown in figure 2.12.

In deep inelastic scattering, one must reconstruct the event kinematics (x, y and Q^2) from the measured final state. On the parton level and in lowest order, this final state consists in first order of a scattered lepton (e or ν) and a scattered quark. These are characterized by an energy and a scattering angle. We use the notation from [24], as described in the table below, and shown in figure 2.13. Polar angles are measured with respect to the z axis, i.e. the proton direction.

E	Energy of the final electron
θ	Polar angle of the final electron
F	Energy of the outgoing (struck) quark
γ	Polar angle of the outgoing (struck) quark
A	Energy of the initial electron (26.7 GeV)
P	Energy of the initial proton (820.0 GeV)

We can reconstruct the kinematics from any combination of the variables E, F, θ and γ. A convenient way of visualizing this, is shown in figure 2.14, where isolines

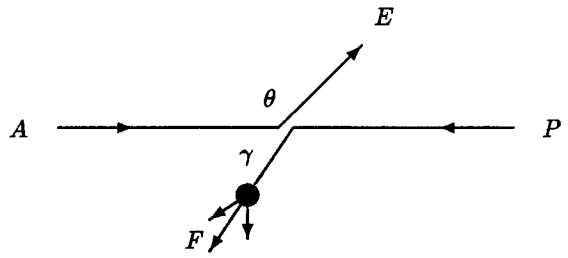

Figure 2.13: *Definition of variables*

of constant E, F, θ and γ are plotted in (x, Q^2) phase space. From this figure one can read the angles and directions of scattered lepton and quark when x and Q^2 are known. For a nice idea of HERA DIS event signatures, one should fold in the cross sections as shown in figures 2.5 and 2.6.

We will now describe in some detail the characteristics of these plots, and methods to reconstruct the kinematics.

2.7.1 Using the scattered electron

One of the characteristic features of neutral current DIS events at HERA is a scattered electron in the final state. This gives an easy to recognize signature for triggering and reconstruction. From this final state electron, characterized by energy E and angle θ (with respect to the proton beam), one can reconstruct the kinematics via the following equations [24]:

$$Q^2 = 2AE(1 + \cos\theta) \tag{2.21}$$

$$x = \frac{A}{P}\frac{E(1+\cos\theta)}{2A - E(1-\cos\theta)} \tag{2.22}$$

$$y = 1 - \frac{E}{2A}(1 - \cos\theta) \tag{2.23}$$

From figure 2.14 it can be seen that at low values of Q^2, the angle at which the electron is scattered $(\pi - \theta)$ is small, and, at not too high values of y, the energy of the scattered electron varies only very little around the electron beam energy. It is clear that, except at high y, the scattering angle is essentially determined by Q^2, and very little by x. This means that the requirement for electroproduction to have the scattered electron in the main detector ($\theta < 176°$) is equivalent to a lower cut-off in the measurable area of Q^2 with a value of about $Q^2 > 3 - 4$ GeV^2.

From figure 2.14, we can draw the following conclusions on the characteristics of scattered electrons in neutral current DIS events:

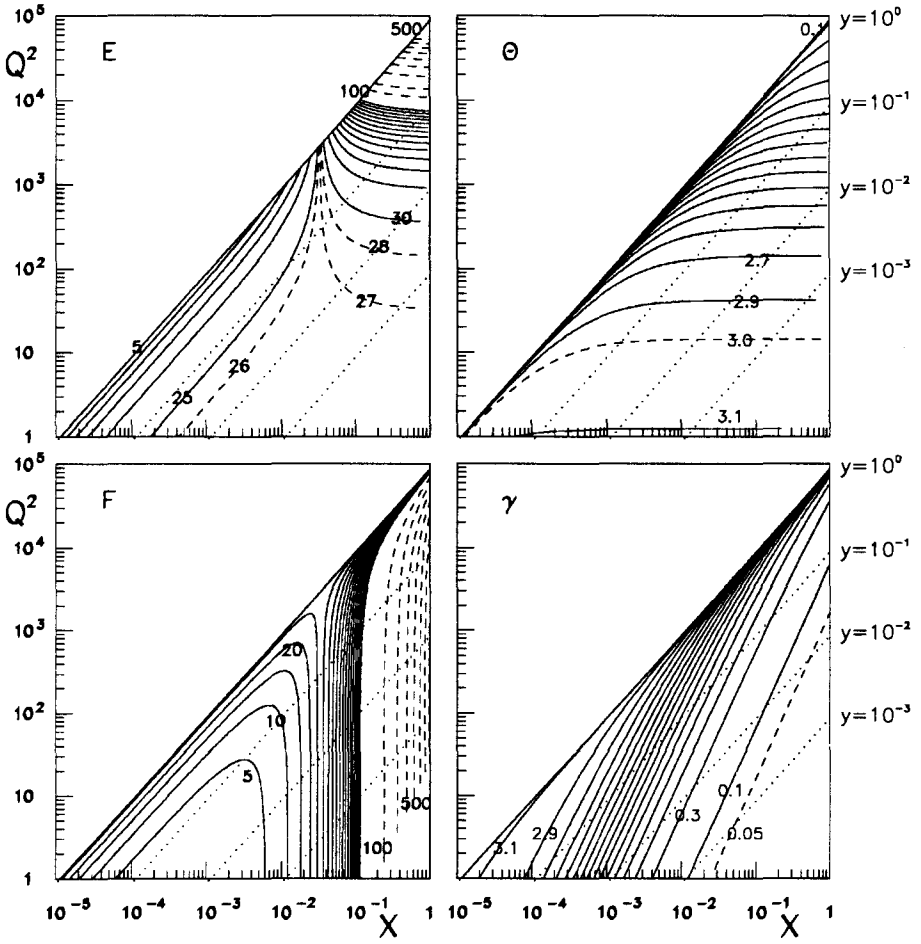

Figure 2.14: *Isolines of constant electron/quark energy and angle in the $x - Q^2$ plane. Upper left: scattered electron energy E in GeV. Upper right: scattered electron angle θ in rad. Lower left: scattered quark energy F in GeV. Lower right: scattered quark angle γ in rad.*

- In a large area of phase space, the scattered electron energy is close to the electron beam energy. Therefore, in the energy spectrum of the scattered electron, one expects a clear peak around the beam energy, which can be useful in calibrating the detector. The differential cross section is such that a lower value of the energy is favoured, but higher energies are also possible at higher x and Q^2;
- Small values of the scattering angle are favoured, which means that in the

majority of the events, the scattered electron will enter the rear part of the detector, to a lesser extent the barrel part, and only occasionally the forward part. The scattering angle increases with Q^2, and is largely independent of x, except at high values of y;

- Many scattered electrons will not hit the tracking detectors. For the ZEUS detector, which will be described in the next chapter, the trackers only see electrons at $Q^2 > 22$ GeV^2 (or even 55 GeV^2 in the case of the central tracking detector). Therefore, in events at low Q^2, the electron can only be measured in the calorimeter.

From figure 2.14 it follows that a measurement of the electron energy contains little information on Q^2 at low Q^2, and no information on x at high Q^2. The angle θ gives a good measurement of Q^2, but not of x. To be more quantitative, one can calculate the following partial derivatives [25]:

$$\left.\frac{\partial Q^2}{Q^2}\right|_E = \frac{dE}{E} \tag{2.24}$$

$$\left.\frac{\partial Q^2}{Q^2}\right|_\theta = -\tan(\frac{\theta}{2})d\theta \tag{2.25}$$

$$\left.\frac{\partial x}{x}\right|_E = \frac{1}{y}\frac{dE}{E} \tag{2.26}$$

$$\left.\frac{\partial x}{x}\right|_\theta = \tan(\frac{\theta}{2})(x\frac{P}{A} - 1)d\theta \tag{2.27}$$

from which one can conclude that the resolution in Q^2 as obtained by the electron variables is very good (except for very small scattering angles), but only if one is experimentally able to obtain an optimum angle reconstruction. However, the resolution in x becomes very poor at low y due to the limited energy resolution. It has been suggested that the electron variables can be used for kinematics reconstruction in the range $5 \cdot 10^{-5} < x < 0.6$, $Q^2 > 5$ GeV^2 and $y > 0.1$ [25].

2.7.2 Using the hadronic energy flow

At HERA, one will be able to obtain a completely independent reconstruction of the kinematics using the hadronic energy flow. In neutral current events, this will bring extra redundancy and an extension of the measurable part of the x-Q^2 plane. For charged current events, the measurement of the hadronic final state provides the only way to reconstruct the kinematics.

From the scattered quark energy F and the angle with respect to the proton direction γ, one can reconstruct the kinematics using:

$$y = \frac{F(1 - \cos\gamma)}{2A} \tag{2.28}$$

$$Q^2 = \frac{F^2 \sin^2\gamma}{1-y} \tag{2.29}$$

$$x = \frac{Q^2}{sy} \tag{2.30}$$

However, one does not directly measure F and γ. One might try to look for jets in the final state that can be associated with a struck quark. For any given event, the ensemble of reconstructed jets may include such a current jet, but also jets originating from hard and soft QCD processes, and a jet originating from the proton (target) remnant. This greatly obscures the direct use of F and γ for kinematics reconstruction.

Instead, the hadronic energy flow as measured in the cells of a segmented calorimeter, can be used without jet finding. Jacquet and Blondel [26] have proposed such a method, based on appropriate weighting of the energy deposits in the cells. The idea is that the hadronic final state consists of a component that has taken part in the hard scattering, and which therefore contains the information on F and γ, and a component that was only a spectator in the interaction, the target remnant. When an expression can be found for x, y and Q^2 where the influence of the latter component is suppressed with respect to the first component, the kinematics can be accurately reconstructed from the hadronic energy flow. The spectator component is characterized by a boost in the original proton direction. The Jacquet-Blondel (JB) method now finds the kinematics via:

$$y = \frac{\sum_i (E_{h,i} - p_{z,i})}{2A} \tag{2.31}$$

$$Q^2 = \frac{(\sum_i p_{T,i})^2}{1-y} \tag{2.32}$$

and again $x = Q^2/sy$. In these equations, the summation formally runs over all final state measured particles, except the scattered lepton. In practice, it is sufficient to sum all calorimeter cells, except those that have been assigned to a final state scattered lepton. In the JB equations, E_h denotes the particle energy, p_z its momentum component in the z direction, and p_T its transverse momentum (to be added vectorially in the summation). As can be seen from the equations, the influence of the target remnant and any hadronic energy lost in the forward beam hole (!), characterized by low p_T and low $E - p_z$, is suppressed by the weighting.

An analysis of the attainable accuracy learns that the JB method is useful in the area $0.01 < x < 0.5$, $Q^2 > 100$ GeV^2, and $y > 0.03$ [25]. The method obviously also works for charged current interactions. A problem with the use of the hadronic energy flow lies in the limited hadronic energy resolution of a calorimeter.

2.7.3 Using mixed methods

Of course, there are more possible combinations of E, F, θ and γ than just (E, θ) and (F, γ). The various methods combining information from the electron and

hadronic energy flow are known as mixed methods. Taking into account the strong points of the electron and JB methods, it is easy to define the mixed variable method:

$$y_{mix} = y_{JB} \tag{2.33}$$

$$Q^2_{mix} = Q^2_{electron} \tag{2.34}$$

$$x_{mix} = \frac{Q^2_{electron}}{s y_{JB}} \tag{2.35}$$

In addition, it has proven to be useful to use θ and γ in the event kinematics reconstruction:

$$y = \frac{\sin\theta(1-\cos\gamma)}{\sin\gamma + \sin\theta - \sin(\theta+\gamma)} \tag{2.36}$$

$$Q^2 = 4A^2 \frac{\sin\gamma(1+\cos\theta)}{\sin\gamma + \sin\theta - \sin(\theta+\gamma)} \tag{2.37}$$

$$x = \frac{A}{P} \frac{\sin\gamma + \sin\theta + \sin(\theta+\gamma)}{\sin\gamma + \sin\theta - \sin(\theta+\gamma)} \tag{2.38}$$

a method which has become known as the 'double angle formula' [24]. Angles are experimentally easier to measure than energies, and do not suffer (in first order) from calorimeter energy calibration problems. Of course, the main problem in the double angle formula lies in the determination of γ.

It has been suggested to calculate γ from the hadronic final state by inverting the Jacquet-Blondel equations [24]. One gets:

$$\cos\gamma = \frac{Q^2_{JB}(1-y_{JB}) - 4A^2 y^2_{JB}}{Q^2_{JB}(1-y_{JB}) + 4A^2 y^2_{JB}} \tag{2.39}$$

Using equation 2.31 and 2.32, this is equivalent to:

$$\cos\gamma = \frac{(\sum p_x)^2 + (\sum p_y)^2 - (\sum(E-p_z))^2}{(\sum p_x)^2 + (\sum p_y)^2 + (\sum(E-p_z))^2} \tag{2.40}$$

When the p_T of the spectator jet and the jet masses are neglected, this is equivalent to the angle of the current jet:

$$\cos\gamma = \frac{\sum_c E_i \cos\gamma_i}{\sum_c E_i} \tag{2.41}$$

where the sum runs over all particles in the current jet.

Chapter 3

The HERA *ep* collider

In this chapter we will shortly review the layout properties of the HERA *ep* collider itself [27]. We will discuss how the luminosity can be determined, and we will talk about beam induced backgrounds for physics.

3.1 The Machine

HERA is an electron-proton (*ep*) collider located on, and around, the site of DESY in Hamburg, Germany. The tunnel in which it resides is approximately a circle, with a circumference of 6336 meters, and a diameter of 5.6 m. It is located 20 to 40 meters under ground level, enclosing the Hamburger Volkspark. Inside the tunnel, two separate synchrotron rings, one for the electron beam and one for the proton beam, form the actual machine. When operating, bunches of electrons and protons circulate in the rings in opposite directions, with the chance of colliding in two interaction points, one in the North Hall and one in the South Hall. Part of the existing system of accelerators at DESY is used for injecting electrons and protons into HERA, at a momentum of 12 and 40 GeV/c respectively. The lay-out of this accelerator system, including HERA, is shown in figure 3.1. This figure also shows a cross section through the tunnel, showing the location of the proton ring on top of the electron ring in the arcs of the machine.

The HERA design envisages a nominal beam momentum of 820 GeV/c for the protons and 30 GeV/c for the electrons, but HERA has operated in 1992 with 820 on 26.7 GeV/c as will be described in more detail in section 8.1. The main design parameters of HERA are given in table 3.1.

Two general purpose detectors have been built for the measurement of particles produced in the *ep* collisions; they are installed and are taking data. The North Hall houses the H1 detector, while the ZEUS detector occupies the South Hall. The work described in this thesis was performed within the ZEUS collaboration. The ZEUS detector will be described in detail in chapter 4.

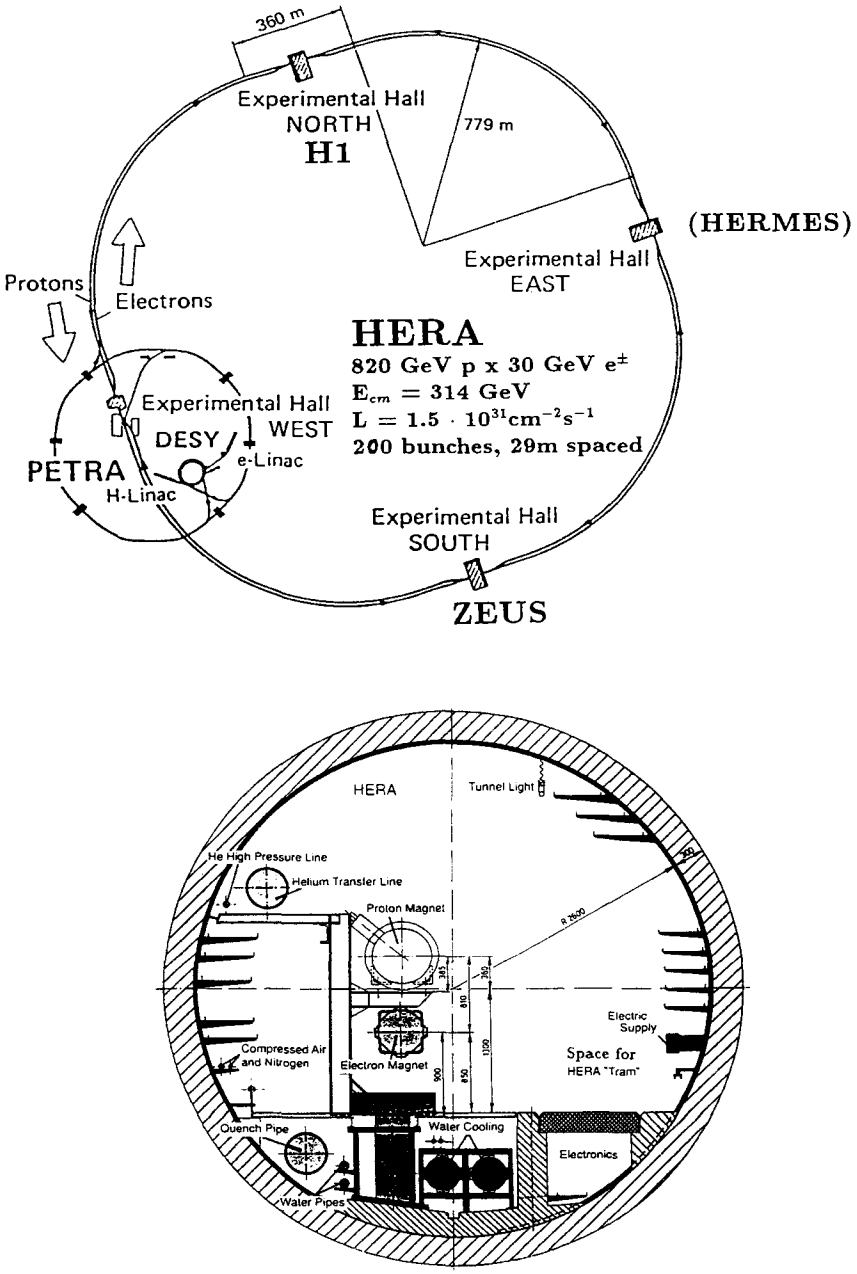

Figure 3.1: *Layout of the DESY accelerator system, including HERA, and a cross section through the HERA tunnel. The location of the experimental halls is also shown in the top figure.*

Parameter	proton ring		electron ring	units
Nominal energy	820		30	GeV
c.m. energy		314		GeV
Q^2_{max}		98400		$(\mathrm{GeV}/c)^2$
Luminosity		$1.5 \cdot 10^{31}$		$\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Number of interaction points		4		
Crossing angle		0		mrad
Free space for experiments		± 5.5		m
Circumference		6336		m
Magnetic field	4.65		0.165	T
Energy range	300–820		12–33	GeV
Injection energy	40		12	GeV
Circulating currents	160		58	mA
Total number of particles	$2.1 \cdot 10^{13}$		$0.8 \cdot 10^{13}$	
Number of bunches		200		
Number of buckets		220		
Time between crossings		96		ns
Filling time	20		15	min.

Table 3.1: *Some design parameters of the HERA ep collider.*

The interesting physics that can be studied in *ep* colliding beam experiments has been described in the previous chapter. The typical expected cross section for hard ($Q^2 > 100$ $(\mathrm{GeV}/c)^2$) NC interactions is 6 nb, for CC interactions 65 pb, and for hard photoproduction a few μb. A complete extraction and analysis of interesting structure functions demands a total event sample corresponding to an integrated luminosity of $\mathcal{O}(500)$ pb^{-1}. HERA has set as its design goal that an integrated luminosity of 100 pb^{-1} per year must be possible, this would mean an instantaneous luminosity of $1.5 \cdot 10^{31}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In order to reach this, HERA operates with a total number of 220 *buckets*, spaced 96 ns apart, that can be filled with groups of particles (*bunches*). It is planned to fill eventually 200 of these buckets. For the experiments this means that they must be ready to deal with a time between bunch crossings of only 96 ns, more than 230 times smaller than at LEP, and only 4 times larger than planned for LHC. What this means for the detector read out and trigger will be discussed in chapter 4.

Reaching a proton momentum of 820 GeV/c in a ring of HERA size, demands very strong magnetic fields in the bending dipoles in the arcs of the machine. This has been achieved by making those magnets in the arcs superconducting. Persistent currents have proven not to be a serious problem [28].

The beams collide head on, i.e. at an angle of 0°, in the interaction points, by steering them with magnets into a common beam pipe in the detector device. The free space for the experiments on both sides of the interaction point is some 6 meters.

The emission of synchrotron radiation naturally polarizes the beams. This ef-

fect is negligible for protons, due to their large mass, but is very important for the electron beam. Theoretically, 80% polarization must be reachable, in 1992 tests more than 55% has been obtained. In order to do experiments with the polarized beams, the polarization axis of the electrons must be rotated parallel or antiparallel to the electron direction (the natural polarization via the emission of synchrotron radiation leads to a polarization axis orthogonal to the plane in which the electrons move), using spin rotators. Apart from the unfolding of structure functions with polarized lepton beams, polarized beams open up the possibility of measuring polarized structure functions. For this purpose, also a polarized target must be available. The HERMES collaboration [29] proposes to use an internal gas target in a storage cell to do so, and intends to build a detector in the East Hall.

HERA also offers the possibility to run with positrons instead of electrons. It is likely that also deuterons can be accelerated in the proton beam line.

3.2 Luminosity Measurement

Any process that is to be used to monitor the luminosity seen by an experiment, must fulfill the requirements that its cross section must be fairly large, so that a short measurement time is needed for a reasonably accurate estimate of the instantaneous luminosity, and that that cross section must be very well known, i.e. calculable and be short of uncertainties. In addition, the process must be experimentally measurable.

For HERA, the *ep*-bremsstrahlung process

$$ep \rightarrow e'p\gamma \tag{3.1}$$

can serve very well for luminosity measurements. For the experimental measurement of this process, the outgoing electron and photon should be measured in coincidence in the so called luminosity monitor (LUMI), with $E_{e'} + E_\gamma = E_e$, i.e. the scattered electron and photon energies must add up to the electron beam energy.

The cross section for this process is given semi-classically by the Bethe-Heitler formula [30]. It has been shown that, within the experimental conditions, this cross section agrees excellently with QED calculations [31]. Also the radiative corrections to this process have been calculated [31], they amount to -0.3% within the measurable area of phase space. It is clear that this cross section has small theoretical uncertainties.

The ZEUS luminosity monitor consists of two separate detectors, one for the photon branch and one for the electron branch. The layout of the luminosity monitor components is shown in figure 3.2.

Photons emerging from the bremsstrahlung process at scattering angles $\theta_\gamma \leq 0.5$ mrad leave the proton beam pipe through a window located at a distance of 92.5 m from the interaction point (IP), and are measured in a lead-scintillator electromagnetic calorimeter at 107 m from the IP. In front of the photon calorimeter, a 2 radiation length carbon filter absorbs the large flux of synchrotron radiation photons,

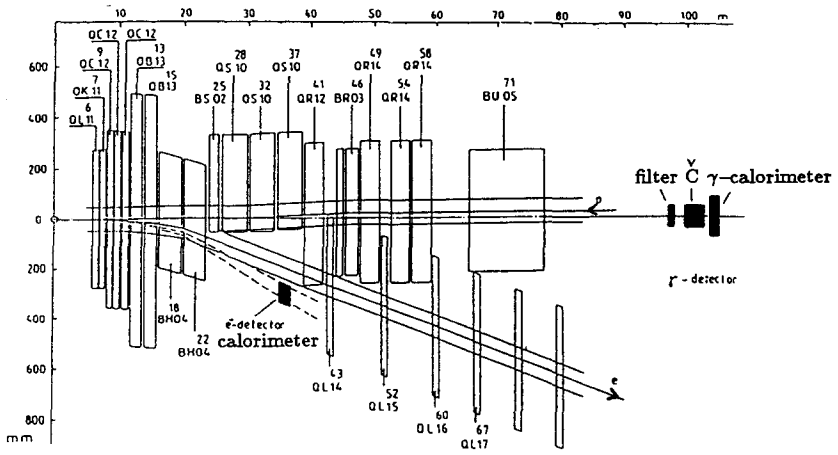

Figure 3.2: *Positions of the luminosity monitor detectors with respect to the electron and proton beam line elements.*

which typically have a low energy in the keV range, whereas the *ep*-bremsstrahlung photons used for the luminosity measurement have energies well above 1 GeV. a Čerenkov counter placed between filter and calorimeter detects events in which the photon has been converted in the filter. The acceptance for bremsstrahlung photons is about 98%, independent of energy.

Electrons scattered at angles $\theta_{e'} \leq 6$ mrad, and with energies $0.2E_e \leq E_{e'} \leq 0.9E_e$ are deflected by the beam magnets away from the nominal electron beam orbit, leave the electron beam line through a window at 27.3 m from the IP, and hit the electron calorimeter located at 34.7 m from the IP. Also this calorimeter is a lead-scintillator sandwich calorimeter. The acceptance is measured to be above 70% and flat for scattered electrons with energies in the range $0.35E_e \leq E_{e'} \leq 0.65E_e$. This region is therefore used in the luminosity measurement.

In addition, the electron calorimeter of the luminosity monitor can be used to tag electrons scattered at very small scattering angles. In this case, no coincidence with the photon calorimeter is required. This electron tagger mode covers the Q^2 range between 10^{-7} and $2 \cdot 10^{-2}$ $(\mathrm{GeV}/c)^2$, although the location of this calorimeter is such that only a fraction of the scattered electrons from events in this range of Q^2 are actually tagged.

The photon and electron calorimeters are calibrated in situ using the observed end point value of the bremsstrahlung photon energy spectrum end fitting the distribution of the sum of electron and photon energies.

The luminosity is easily obtained from the observed *ep*-bremsstrahlung rate R_{ep} by $L = R_{ep}/\sigma_{obs}$, where σ_{obs} is the *ep*-bremsstrahlung cross section corrected for de-

tector inefficiencies and acceptances. The uncertainty in σ_{obs} is small, therefore the error in the luminosity measurement is largely governed by the uncertainty in R_{ep}. The luminosity measurement suffers from a number of background processes. By requiring a coincidence between a high energy electron and photon, this background is dominated by interactions of electrons with the remaining gas in the beam pipe; these will be described in the next section. The resulting systematic uncertainty in the luminosity measurement is approximately 5% for the 1992 data.

3.3 Beam Induced Backgrounds

The backgrounds for experimenting at HERA induced by the beams, can be separated in backgrounds from the electron beam, and the potentially much more dangerous backgrounds from the proton beam.

Electron Beam

The electron beam produces a large amount of synchrotron radiation. Most of this, however, is produced in the bending arcs far away from the interaction point. The remaining radiation can be effectively shielded with a set of masks and collimators in the beam line.

Electrons can also produce bremsstrahlung under influence of nuclei of remaining gas in the beam pipe. Such electron-gas events can deposit a large amount of energy in the rear part of the main detector, which will trigger the trigger system to accept the event. However, these events give very little energy in the forward part of the detector, in contrast to real *ep* events, and can thus be removed in the subsequent offline analysis. Electron-gas events form a large background for the luminosity measurement when the electron, or the photon, or both, are detected by the luminosity monitor. The contribution of this background to the total counting rate of the luminosity monitor can be estimated by including in the machine a *pilot* electron bunch, that does not collide with a proton bunch by leaving the corresponding proton bucket empty (see section 8.1). Let the totally measured counting rate of the luminosity detector be R_{tot}, the rate from the pilot bunch be R_{pilot}, the total current in the electron ring be I_{tot}, and the current in the pilot bunch be I_{pilot}, then the actual *ep*-bremsstrahlung rate R_{ep} for luminosity measurement can be estimated as:

$$R_{ep} = R_{tot} - R_{pilot}\frac{I_{tot}}{I_{pilot}} \tag{3.2}$$

The contribution of electron–gas interactions can easily exceed the actual *ep*-bremsstrahlung rate at low luminosities, leading to a relatively large uncertainty in the luminosity measurement.

Proton Beam

The proton beam is accompanied by a fairly large number of halo muons (beam muons), which are caused by decaying pions, created in collisions of beam particles with the rest gas or the beam pipe wall, magnets or collimators. These muons do not form a problem in the analysis, and can be used for calibration purposes of the calorimeter.

Events caused by collisions of particles in the proton beam with remaining rest gas in the beam pipe form the major source of background, since the total proton-proton cross section is very high. Although the p_T in such collisions is generally low, the huge cross section causes these processes to form a large background for physics. The rate of events depends on the proton flux and the pressure in the beam pipe, and has been estimated to be around 0.5 kHz/m for design conditions. For the actual rate seen by the detector, this rate has to be integrated over a large distance, since events occurring far upstream from the interaction point can still cause activity in the main detector. The integral is obtained by convoluting with a number denoting the fraction of events actually causing enough activity in the main detector to activate a trigger. It has been estimated that this beam gas background rate may be as high as $\mathcal{O}(50)$ kHz, which is 5 orders of magnitude larger than the rate of interesting physics. In chapter 4 we will explain how this background can be suppressed with a multi-level trigger. Figure 3.3 shows a typical example of a beam-gas event occurring upstream of the ZEUS detector. The layout of the ZEUS detector will be explained in the next chapter; the figure shows an event display picture of the ZEUS detector, in particular the calorimeter and tracking detectors left, and the veto wall right, the hits in the tracking chambers, the calorimeter and the veto wall, and the reconstructed tracks and clusters. It is clear that the event is characterized by a high number of tracks with an apparent upstream vertex, hits in the veto wall, and a relatively large energy deposit in the rear calorimeter.

A final class of proton beam induced background events are those caused by collisions of beam halo particles with the beam pipe wall, magnets, collimators or shielding, which cause a spray of secondary particles in the forward direction. They can be treated in the same way as beam gas events.

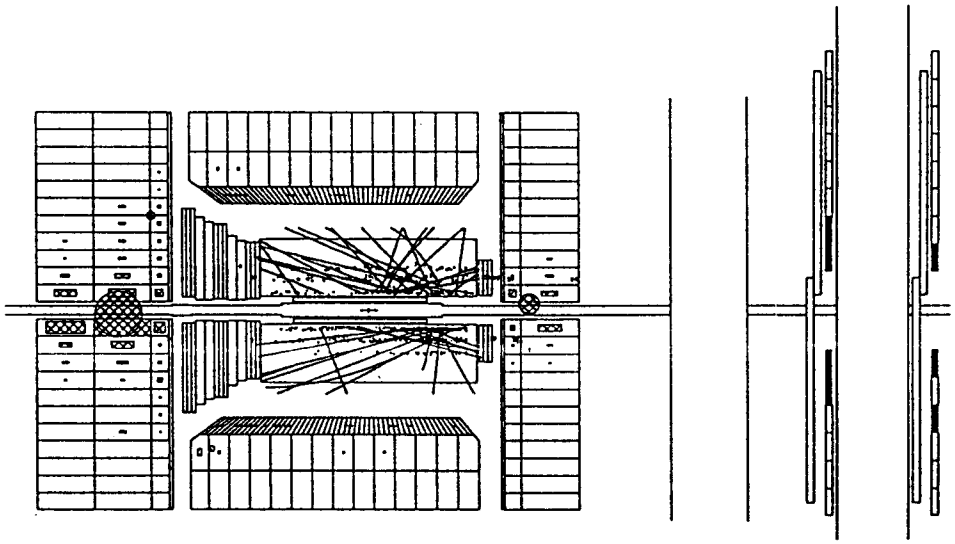

Figure 3.3: *A typical beam-gas event, occurring upstream in the proton beam line, as measured in ZEUS.*

Chapter 4

The ZEUS detector

4.1 Design Considerations

As in every experiment, the criteria for a detector follow from the conditions of the accelerator and the physics that one wants to be able to measure. The various interesting physics processes at HERA, and their experimental signature, have been described in chapter 2. From these signatures follow the following detector design criteria:

- The asymmetry between the electron and proton beam momenta boosts the center of mass system in the forward direction, with respect to the laboratory frame. This asks for an asymmetric detector.
- NC DIS events have a scattered electron in the final state, whose kinematics has already been described in section 2.7. The detector must be able to measure isolated electrons. For DIS events, these typically have energies exceeding 5 GeV, but in events in which a heavy quark decays semileptonically, the energy of the produced electron is typically $\mathcal{O}(1)$ GeV. The detector thus needs a good tracking system and a good electromagnetic energy resolution.
- CC DIS events are characterized by large missing transverse momentum. The measurement of this process requires the detector to be hermetic, i.e. cover as much of 4π solid angle as possible. The same argument holds for the measurement of many exotic processes. In addition, the hadronic final state provides an independent, and sometimes more accurate, handle on the reconstruction of the kinematics of NC DIS events than the electron, and is the only possibility for CC DIS events. This means further that the hadronic energy resolution of the detector should be as good as possible. The segmentation should be sufficient to provide a good jet reconstruction capability.
- Measuring the momentum of charged tracks requires a high magnetic field, but the magnet should be thin in order not to disturb the calorimetric en-

ergy measurement, when placed inside the calorimeter. Such a placement is necessary if photomultipliers are used in the calorimeter readout.

- The detector should be efficient for the detection and identification of charged leptons like electrons and muons in the vicinity of hadron jets, such as can be created in semileptonic decays of heavy quarks, or in exotic processes. This requires an excellent electron-pion separation capability, which can be obtained with a combination of tracking and calorimetry, and a good muon measurement system.
- A good tracking detector and vertex detector enable the accurate reconstruction of primary and secondary vertices, as well as the exclusive reconstruction of particles like D^*, that indicate heavy quark production.
- For the measurement of the luminosity, and for measuring very forward scattered electrons in photoproduction events, a small angle electron and photon tagging system is required.
- Very forward scattered protons could be measured with a very forward proton spectrometer system.
- The detector should be able to work in HERA conditions, which means a very short bunch crossing time, a high background interaction rate, and a high radiation level. This requires fast detectors with good time resolution, pipelining technology, and an excellent trigger system.

It has been shown that a good measurement of DIS events is greatly enhanced by good hadron calorimetry [32, 33]. The ZEUS detector implements this in the uranium/scintillator calorimeter (CAL), which in addition has a good electron/pion separation capability using layers of silicon diodes as hadron–electron separator (HES) built inside. The calorimeter is fast, and has a good time resolution.

The ZEUS detector is located in the HERA south hall and measures approximately 20 by 11 by 12 meters, and has a total weight of 3600 tonnes. In addition, two components are located in the beam line.

In the following sections we will describe the detector components, with special emphasis on the high resolution calorimeter. A full and detailed overview can be found in [34, 35].

4.2 Detector Components

A cross section through the ZEUS detector is shown in figures 4.1 and 4.2. Note that the asymmetry in the beam momenta is reflected in the detector layout. The coordinate system is defined in figure 2.12. The positive z-axis, which corresponds to the proton direction, defines the forward hemisphere, the negative z-axis the rear hemisphere.

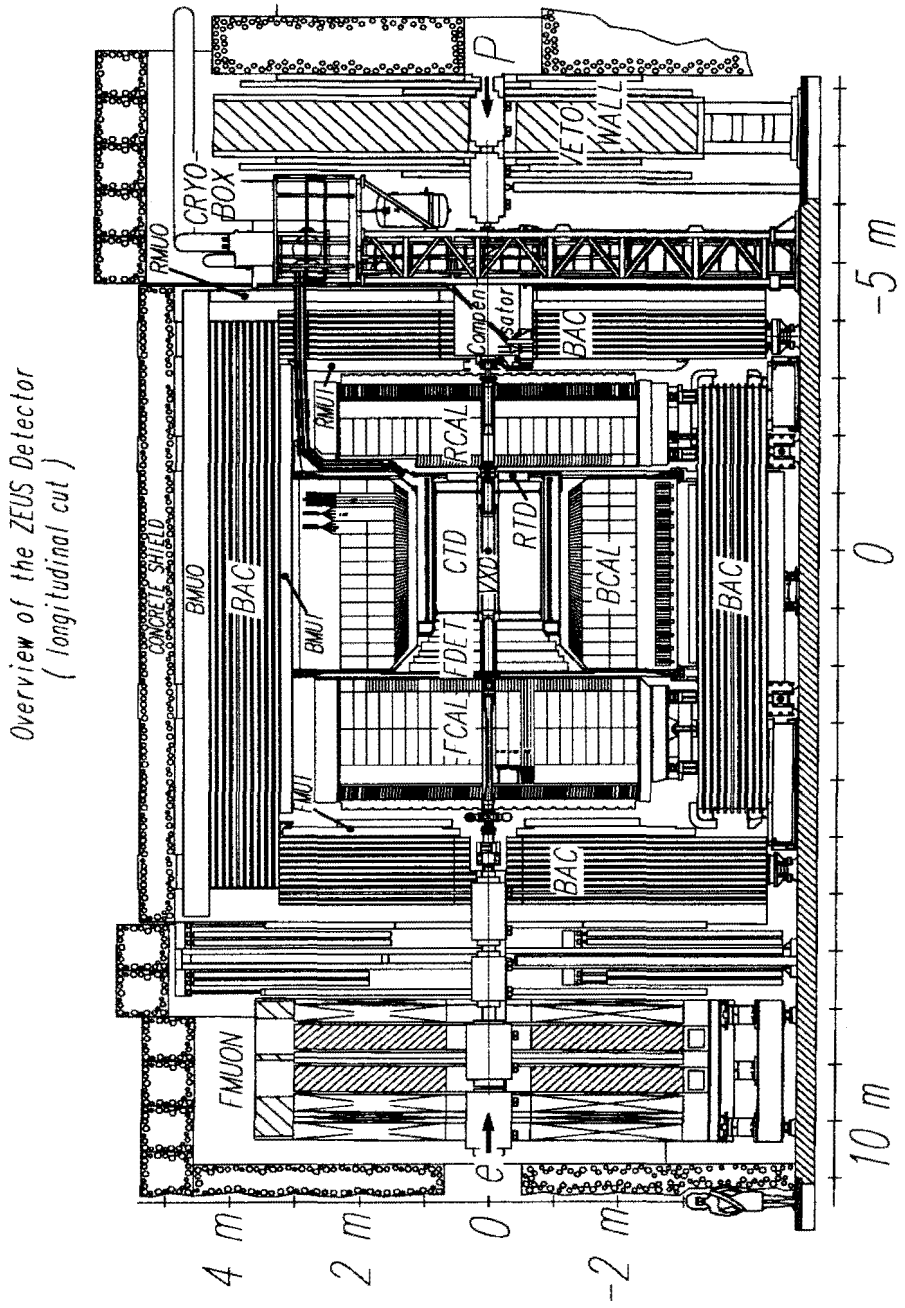

Figure 4.1: *A cross section through the ZEUS detector along the beam*

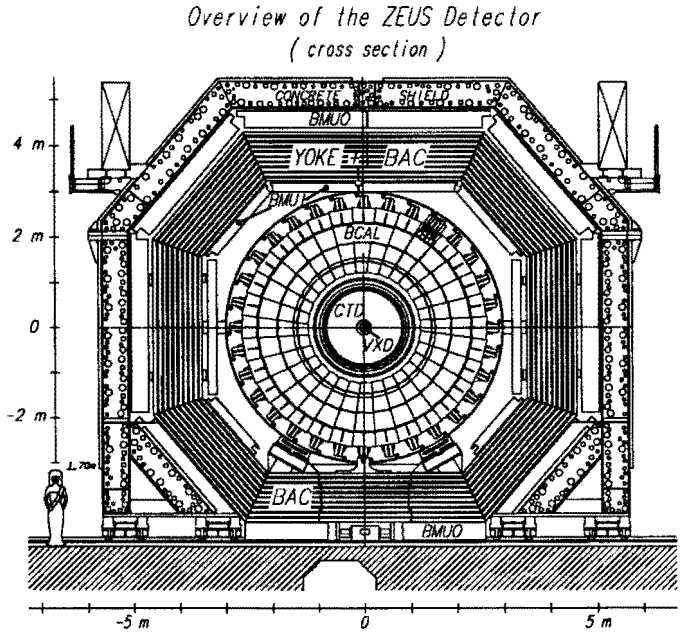

Figure 4.2: *A cross section through the ZEUS detector perpendicular to the beam*

The detector is described from inside out:

- The interaction region and beam pipe elements. The central part of the beam pipe is made of aluminium, its diameter is approximately 170 mm. The beam pipe in the detector is equipped with masks and collimators against the high background of synchrotron radiation and beam-gas or beam-wall interactions. The **C5** collimator, which is located behind RCAL, is equipped with scintillator planes, whose signals provide information on upstream beam-gas or beam-wall interactions as well as on the timing of the two beams, which is crucial for the rejection of beam-gas events.

- The vertex detector (**VXD**). Its main purpose is to measure the main vertex and to detect short-lived particles by reconstructing secondary vertices, to improve the momentum and angular resolution of charged particles and to help in the pattern recognition. The vertex detector is a combination of a jet chamber and a time-expansion chamber with an inner radius of 100 mm and an outer radius of 156 mm. In the radial direction, 12 sense wires, with a length of 1.6 m, are placed, 3 mm apart and alternated with field wires. The spatial point resolution in the $r-\phi$ plane, which is the direction perpendicular to the beam axis, ranges between 35 and 55 μm, and the double track resolution is 500 μm. This considerably helps the CTD in determining the impact point and increasing the momentum resolution of charged tracks. Combining the

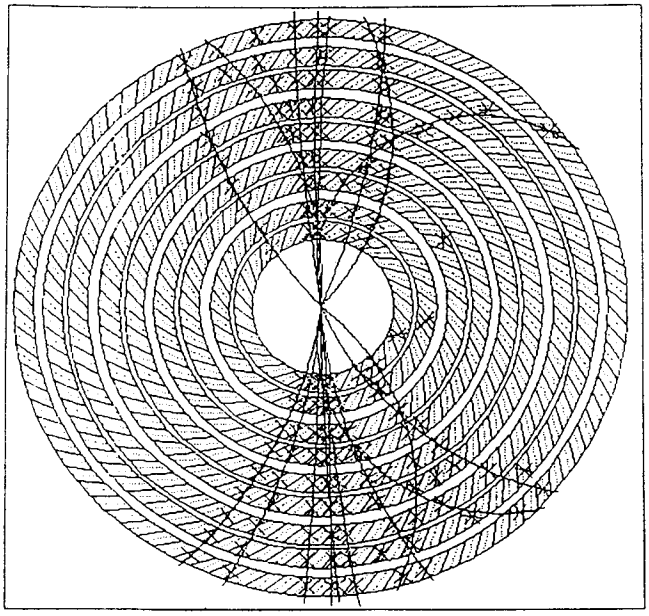

Figure 4.3: *A cross section through the CTD, perpendicular to the beams, showing the wire layout. Superimposed are the hits and tracks of a simulated event.*

information of VXD and CTD leads to a vertex resolution in $r-\phi$ between 30 and 80 μm.

- The central tracking detector (**CTD**). This is a cylindrical wire chamber with nine superlayers of eight sense wire layers each, as shown in figure 4.3. Five of the superlayers have wires parallel to the chamber axis, four layers are stereo layers which have wires with a small angle (5° or 7°) with respect to the beam, and which are used to determine the z coordinate of the hits. The inner radius of the chamber is 16.2 cm, the outer radius 85 cm, and the length is 241 cm. The chamber uses a 50/50 argon/ethane mixture bubbled through ethanol. The position resolution in the $r-\phi$ plane is 100-120 μm (depending slightly on polar angle θ) when the full $r-\phi$ electronics are present. The z resolution using the stereo wires equals 1.0 to 1.4 mm, and when using z-by-timing 4 cm. The momentum resolution of the CTD with full readout equals $\sigma(p)/p = 0.0021p \oplus 0.0029$ for a particle traversing all 9 superlayers. The accuracy is lower for particles that only hit part of the supercylinders. The detector measures dE/dx with an accuracy of better than 6 %, which is useful in particle identification. The CTD covers an angular range of $15° < \theta < 164°$.

- Additional chambers which are located in the forward and rear direction: the forward and rear tracking detectors (**FTD** and **RTD**). These chambers

are vital in measuring the forward and backward tracks that hit one or no layer of the CTD. At HERA, such tracks are numerous, and the forward and backward chambers considerably increase the tracking performance in these regions, where the track density is high. The RTD is a single planar chamber of the same type as used in the FTD. The FTD consists of three such chambers, each consecutive chamber in z has a larger diameter. Each chamber consists of three layers of drift cells with fixed wire orientations of 0°, +60° and −60°. The used gas is identical to the CTD gas, the spatial resolution equals 120-130 μm, the two track resolution is 2 mm. The whole tracking system covers polar angles between 7.5° and 170°, of which FTD covers $7.5° < \theta < 28°$ and RTD $160° < \theta < 170°$.

Between the 3 FTD chambers, modules of a transition radiation detector (**TRD**) are positioned. Each of the 4 modules of the TRD consists of a radiator stack (polypropylene fibres) followed by a drift chamber. The TRD will be used for electron identification in the momentum range from 1 to 30 GeV in the forward direction. FTD and TRD together form the forward detector FDET.

- The magnet coil. A superconducting solenoid of inner radius 92.5 cm and length 246 cm, operated at a current of 5000 A, provides a magnetic field of 1.8 T for the inner detectors. The coil has an aluminium support structure, with a NbTi/Cu superconducting cable. The coil has a thickness corresponding to 0.9 radiation lengths (X_0). The influence of the field on the beams is corrected for by a compensating superconducting solenoid, located behind the RCAL.

- The high resolution uranium calorimeter (**CAL**). This detector component will be extensively described in section 4.4. The CAL measures the energy of most particles, charged and neutral, by absorbing them. Only muons, neutrinos, and particles produced in showers at the back of the calorimeter may leave the CAL at the rear.

- The hadron electron separator (**HES**). Mounted inside the calorimeter, several single layers of silicon diodes located after a few radiation lengths will be useful in the separation of electrons (or photons) and hadrons, by using the differences between electromagnetic and hadronic shower properties, as will be explained in section 4.3. More details of the HES are described in section 4.4.

- The return yoke and backing calorimeter (**BAC**). The iron yoke for the return of the magnetic flux has the shape of an octogonal cylinder which is closed by endcaps on both sides. Mechanically, the yoke splits into three parts: two clam shells that can be opened, and a bottom yoke. The total dimensions are 1037 by 910 by 860 cm, the total weight is 1962 tonnes. The yoke can be magnetized by copper coils, in order to create a magnetic field for an independent measurement of the momentum of muons in the barrel muon chambers. It consists of slabs of iron separated by gaps instrumented with proportional

tubes, and as such forms a backing calorimeter. This calorimeter measures the energy leakage out of the main calorimeter, as well as the trajectories of muons.

- The barrel and rear muon detector (**BMUON, RMUON**). This detector component consists of muon chambers at the inside and outside of the iron yoke, and measures tracks penetrating the calorimeter, and cosmic rays. By measuring the momenta of reconstructed track segments, and comparing this measurement to matching track segments in the CTD, prompt muons can be detected. Each barrel chamber, with a typical size of 3 by 8 meters, consists of 2 double layers of limited streamer tubes, with wires oriented along the beam axis, and is equipped with external readout strips orthogonal to the wires. The position resolution is better than 1 mm. The barrel muon chambers accept muon tracks at $\theta > 34°$. In order to reach the muon chambers, muons must traverse the uranium and backing calorimeters. Since a minimum ionizing particle typically deposits $\mathcal{O}(2)$ GeV in the uranium calorimeter, this is the minimum energy that is needed for the muon in order to be detected in the inner muon chamber. In crossing the iron yoke between inner and outer muon chambers, the muon loses approximately an other 1 GeV of energy.

- The forward muon detector (**FMUON**). Following the requirements posed by the HERA asymmetry, the forward direction is equipped with special muon chambers, which give a completely independent measurement of the muon momentum up to at least 100 GeV down to very low angles. The detector consists of a toroidally magnetized iron region, interleaved with sections of drift chambers, limited streamer tubes and time of flight counters. The outer diameter of the toroids is 6 m, the average magnetic field inside is 1.7 T. The forward muon detector is mounted on its own rails, separate from the main detector.

- The veto wall (**VETO**). The veto wall is an iron wall, equipped with two layers of scintillator hodoscopes on both sides, and is placed 7 m from the interaction point in the electron direction. Its dimensions are 800 by 760 by 87 cm, with a square hole of 95×95 cm^2 in the center for the beam pipe and magnets. Its main purpose is to shield the detector against particles from the proton beam halo and to give a veto for beam-gas interactions that could be measured in the detector.

  The above elements form the main detector. In addition, one component is placed in the electron beam line, and one in the proton beam line (not shown in the figures).

- The luminosity monitor and small angle electron tagger (**LUMI**). The luminosity measurement and the layout of the luminosity detector have been

described in section 3.2. Additionally, the electron calorimeter can detect electrons at very small angles emerging from physics processes like photoproduction.

- The leading proton spectrometer (**LPS**). The proton debris of *ep* interactions is emitted in the very forward direction, but a subsample containing leading protons can be detected using an array of 6 Roman Pots, located between 24 and 90 m from the interaction point along the proton beam.

The main detector is placed on rails and can as such be moved in and out of the beam line. The major part of the electronics not mounted on the detector itself is located in a three story building located next to the detector (Rucksack), which is connected by drag chains to the detector and can move more or less independently of it. The Rucksack is connected to a computer room in the south hall by fibre optic links, the whole detector is operated from a control room located at the ground floor of the hall. The computer room finally is connected by an optic fibre link to the DESY IBM mainframe computer.

4.3 Calorimetry

In the context of high energy physics experimentation, the term *calorimetry* is used for the measurement of particle energies. A calorimeter is an apparatus that accomplishes this by destructive methods, i.e. absorbing the incoming particles and converting a given fraction of their energy to a measurable signal. The processes occurring in a calorimeter when a high energy particle enters create a very large amount of secondary particles, which in turn give rise to new particles, in a cascade-like way which is known as a *particle shower*. The main properties of the detector are:

- A calorimeter is able to measure the energy of charged as well as neutral particles;
- A calorimeter is fast, typical response times are of the order of 100 ns when scintillator is used as readout medium;
- The development of a particle shower in a calorimeter is a statistical process: the relative accuracy of the energy measurement improves with energy, whereas the needed dimensions of a calorimeter increase only very slightly with increasing energy;
- A segmented readout of the signal allows the reconstruction of the position and direction of incoming particles, and together with the different characteristic shower patterns for various particles, a form of particle identification can be done.

These properties make that calorimeters play an important role in future high energy experiments, which are characterized by high rates, and where tracking detectors need a high magnetic field and lots of space, whereas their resolution deteriorates with increasing energy [1].

For a detailed overview of the complex and exciting field of calorimetry, we refer to [36]. In this section we will highlight the most important points.

It is important to realize that there are large differences in the shower development for various types of particles. Muons behave in first approximation as minimum ionizing particles (*mips*), that penetrate a calorimeter loosing an almost fixed amount of energy that only depends on the type and amount of material traversed, and which may be small compared to their actual energy. The showering properties of muons can not be totally neglected, but are small. The difference between *electromagnetic* and *hadronic* showers is very important and will be described in the following sections.

When the absorbing material acts at the same time as a sensitive (i.e. signal producing) material, the calorimeter is said to be *homogeneous*. Homogeneous calorimeters are often used in the measurement of pure electromagnetic showers (electrons, positrons, photons) and can have a very good energy resolution. However, they do not have enough stopping power for hadronic showers to be fully absorbed in a limited distance. A different and commonly used calorimeter design therefore makes use of alternating layers of heavy, but insensitive, absorber material, and lighter sensitive signal producing material. Such calorimeters are known as *sampling calorimeters*. For the absorber layers, heavy material like iron, lead or uranium are often used; the sensitive material can be a gas, liquid argon (or xenon), plastic scintillator or silicon diodes. We will limit ourselves to these sampling calorimeters, and in particular the combination of uranium and plastic scintillator.

Sampling Calorimetry

In a sampling calorimeter, the ionization energy loss of shower particles traversing active (sensitive) layers is measured; this is called the visible energy. This fraction of the particle energy is usually much smaller than the fraction deposited in the absorber layers, which actually make the incoming particle stop. The *sampling fraction* f_{samp} is defined as:

$$f_{samp} = \frac{E_{visible}}{E_{total}} = \frac{E_{visible}}{E_{invisible} + E_{visible}} \tag{4.1}$$

and depends on the type of particle in the shower (shower component).

It is convenient to normalize these sampling fractions to the sampling fraction

[1] Nevertheless, the information from tracking detectors is so valuable that also they will be very important for future experiments.

for minimum ionizing particles f_{samp}^{mip}. For electrons for example we have:

$$\frac{e}{mip} = \frac{f_{samp}^{electron}}{f_{samp}^{mip}} \tag{4.2}$$

and analogue expressions can be written for charged hadrons (h), neutrons (n), and photons (γ). Various effects that play a role in the development of a shower can raise or lower these ratios in various calorimeter geometries.

The above defined sampling fraction is an average over many particles. The statistical showering processes can make the actual energy deposit in the active layers vary for individual showers or shower parts. The resulting fluctuations induce an uncertainty in the energy measurement that translates into an finite energy resolution of the calorimeter. An important contribution to the uncertainty of the energy measurement in sampling calorimeters is caused by *sampling fluctuations*, i.e. the fluctuations in the number of charged shower particles that traverse active layers and contribute to the signal.

4.3.1 High Energy Electromagnetic Showers

The interactions of incoming electrons, positrons or photons are of pure electromagnetic origin. They are fully described by QED, and nuclear reactions play no role. That makes electromagnetic (EM) showers easy to understand and simulate, though the shower development itself is by no means a simple process. The relevant processes for high energy electrons and positrons are:

- **Bremsstrahlung**: charged leptons radiate photons under the influence of the field surrounding heavy nuclei. These photons have a steeply falling energy spectrum. The energy loss is proportional to Z^2, where Z is the atomic number of the material traversed;
- **Ionization**: charged particles loose energy by ionizing the traversed medium. The energy loss is proportional to $Z\log(Z)$.

whereas photons can undergo the following processes:

- **Pair production**: photons with enough energy can produce an electron-positron pair under the influence of the field of nuclei. The cross section is proportional to Z^2. This is indeed the dominating effect at high photon energies;
- **Compton scattering**: the photons scatter on atomic electrons, the cross section is proportional to Z;
- **Photoelectric effect**: when the photon is absorbed by an atomic electron, the atom can be ionized. The cross section is proportional to $Z^4 - Z^5$, dominating at lower photon energies.

The radiation length X_0 sets the typical scale for electromagnetic showers. For a given material, it is defined by [37]:

$$\frac{1}{X_0} = 4\alpha \frac{N_A}{A} Z^2 r_e^2 \ln(183 Z^{-1/3}) \tag{4.3}$$

with α the electromagnetic coupling constant, A the atomic mass of the material, Z the atomic number of the material, N_A Avogadro's number, and r_e the classical electron radius. A radiation length equals the distance over which the energy of a high energy electron on average drops by a factor $1/e$ purely by Bremsstrahlung:

$$< \frac{dE_{Brems}}{dx} >= -\frac{E}{X_0} \tag{4.4}$$

where x represents the path length of the incoming electron in units of radiation lengths, and E equals its energy. In very good approximation, this equation is independent of the material type. For uranium, $X_0 = 0.32$ cm [38]. The energy loss by ionization of an electron traversing one radiation length defines the critical energy ϵ_c:

$$\frac{dE_{ion}}{dx} = -\frac{\epsilon_c}{X_0} \tag{4.5}$$

with $\epsilon_c \approx 6$ MeV in uranium [39].

Since, theoretically, the production of e^+ e^- pairs by high energy photons is related to Bremsstrahlung of charged leptons, the typical distance scale of this process is related to the radiation length of a material. It can be shown [37] that the mean free path λ_{pair} of a photon for pair production equals $9/7$ X_0.

The development of an EM shower can be described as follows. In the beginning, at high energies, processes like pair production and Bremsstrahlung dominate, strongly increasing the number of particles in the shower. When the average particle energy falls below the critical energy ϵ_c, energy loss through ionization becomes dominant. Below this energy, more particles tend to get absorbed than are created in multiplicative processes, and the number of shower particles decreases until the shower dies out. In this low energy part of the shower, the measurable signal is produced mainly via ionization of the active layers by relatively slow charged particles.

Shower Shape

The shower development and shape is mainly determined in the cascade process at high energies, and can therefore be described in terms of the radiation length X_0 and the Molière radius R_M. The radiation length is a measure of the average free path between interactions, and therefore governs the longitudinal shape of the shower, whose profile can be parametrized by [39]:

$$\frac{dE}{dt} = E \frac{b^{\alpha+1}}{\Gamma(\alpha+1)} t^{\alpha} e^{-bt} \tag{4.6}$$

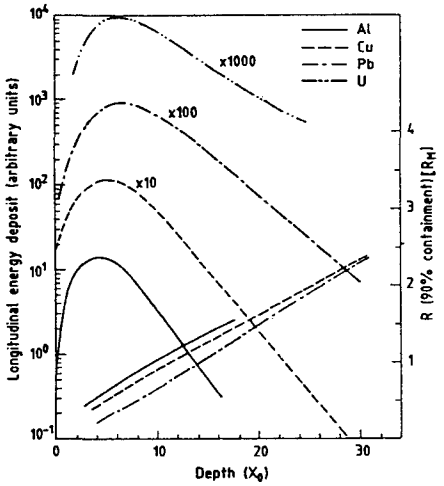

Figure 4.4: *Longitudinal shower development of a 6 GeV electron in Al, Cu, Pb and U, showing the scaling in units of X_0. Also given is the shower radius in units of R_M for 90% shower containment as a function of the shower depth.*

where E represents the primary particle energy, t is the depth measured in units of X_0, $b \approx 0.5$, and $\alpha = b \cdot t_{max}$, where t_{max} equals the position of the shower maximum in units of X_0:

$$t_{max} \approx \ln(E/\epsilon_c) - c \tag{4.7}$$

with $c = 1.1$ for electrons and $c = 0.5$ for photons. An example for the longitudinal shower shape of a 6 GeV electron is given in figure 4.4.

The length needed to contain 98% of an EM shower is phenomenologically given by:

$$L_{98} \approx 3t_{med} = 3(\ln(E/\epsilon_c) + a) \tag{4.8}$$

with $a = 0.4$ for electrons and $a = 1.2$ for photons, and $t_{med} \approx t_{max} + 1.5$ equals the depth, in units of X_0, at which half of the particle energy is deposited. As an example, a uranium scintillator calorimeter with an effective $\epsilon_c = 10.6$ MeV needs to be 25 X_0 deep in order to contain 98% of the shower energy of a 30 GeV electron [32].

The transverse energy distribution is determined by multiple scattering of electrons and traveling Bremsstrahlung photons, and scales with the Molière radius R_M, which describes the average lateral deflection of electrons with an energy ϵ_c after one radiation length X_0:

$$R_M = E_S \frac{X_0}{\epsilon_c} \tag{4.9}$$

with $E_S = 21$ MeV from multiple scattering theory. For uranium, $R_M = 0.96$ cm (for the uranium/scintillator structure used in the ZEUS calorimeter, $R_M = 2$ cm). The lateral containment as a function of depth is given in figure 4.4. 95% of the

total deposited energy is contained in a cylinder with a radius of 2 R_M, and 99% within 3 R_M. The transverse shape can be best parametrized by two superposed exponential functions:

$$\frac{dE}{dr} = A_1 e^{-|r|/B_1} + A_2 e^{-|r|/B_2} \tag{4.10}$$

with the first exponential describing the core, and the second describing the tail of the shower shape, so that $B_1 < B_2$. The energy in the core is larger than in the tail: $A_1 > A_2$.

4.3.2 Hadronic Showers

Incoming hadrons not only undergo electromagnetic, but also nuclear (strong) interactions. The enormous variety of these processes makes a hadronic shower much more complicated. The cascade idea, however, of increasing multiplicity followed by energy loss and dying out remains valid. The secondaries are created mainly in inelastic collisions with nuclei of the absorber material, which tend to give rise to the following processes:

- the production of charged hadrons that loose their energy by ionization, until they undergo new strong interactions;
- the production of neutral hadrons that only undergo strong interactions (or decays), but do not loose their energy in any other way.
- the production of neutrinos that escape undetected. In a finite calorimeter, particles produced in the tail of the shower may also leak out of the detector;
- the production of particles that shower purely electromagnetically, like π^0's or η's, and which give rise to an electromagnetic component in every hadron shower;
- the production of excited nucleons, which fall back to the ground state releasing low energy photons, nucleons or fragments of nuclei (e.g. α particles), or which may undergo spallation or fission.

The involvement of nuclei in hadronic showers leads to the observation that the response of calorimeters to hadronic showers is usually lower than to EM showers. This can largely be explained by the fact that in nuclear collisions in heavy materials, many charged hadrons and nuclear fragments are produced that remain in the absorber material, and do not reach the active layers. In addition, part of the energy goes into nuclear binding energy or escapes undetected in the form of neutrinos. The energy resolution of a hadron calorimeter is usually worse than that of an electromagnetic calorimeter, since in hadronic showers it is mainly determined by the enormous variety of nuclear reactions that can occur. In addition, each hadronic shower has a varying electromagnetic component, caused by the creation in a hadron shower of particles that only shower electromagnetically themselves. What this means for the energy measurement will be discussed in the next section.

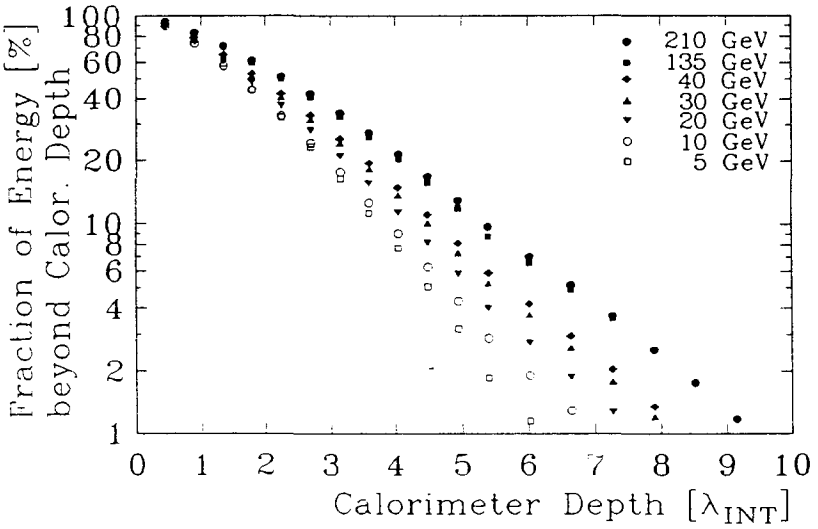

Figure 4.5: *The fraction of energy deposited beyond a given depth, measured in a uranium scintillator calorimeter, for hadron showers of different energies.*

Shower Shape

The most important parameter for the description of the development of a hadronic shower is the *nuclear interaction length* λ_{int}, which can be interpreted as the mean free path between hadronic interactions:

$$\lambda_{int} = \frac{A}{\rho N_A \sigma_i} \tag{4.11}$$

where A is the mass number of the absorber, ρ is the specific density, N_A is Avogadro's number, and σ_i represents the inelastic cross section, which is largely independent of hadron energy and the type of hadron. For uranium, $\lambda_{int} = 10.5$ cm [38].

Figure 4.5 shows an example of the measured energy deposition profile (as fraction of energy deposited beyond a given depth) for hadron showers of different energy in a uranium scintillator calorimeter. The position of the shower maximum, in units of λ_{int} from the entrance face of the calorimeter, is parametrized by [39]:

$$t_{max} = 0.2 \ln(E) + 0.7 \tag{4.12}$$

and the depth needed to contain 95% of the shower energy is given by:

$$L_{95} = t_{max} + 2.5 E^{0.13} \tag{4.13}$$

which implies that 4.5 λ_{int} is needed to contain a 10 GeV hadronic shower, which only increases to 7.0 λ_{int} for a 275 GeV shower.

The transverse shower shape is again best described by a sum of two exponentials, one for the core and one for the tail of the shower. An example of a measured

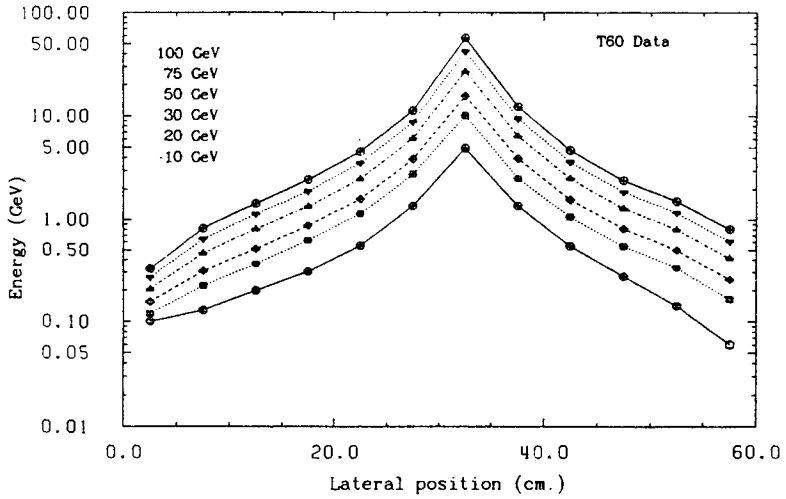

Figure 4.6: *Transverse projections of hadronic shower profiles, integrated over the whole calorimeter depth, as measured in a test calorimeter. The lines are meant to guide the eye.*

transverse shape in a uranium scintillator calorimeter is given in figure 4.6. The transverse shower distribution has a narrow core (0.1 - 0.5 λ_{int} FWHM) that increases with depth. 95% of the shower energy is contained laterally within a radius of 1 λ_{int}.

For a uranium calorimeter, the high values of $\lambda_{int}/X_0 = 33$ and $\lambda_{int}/R_M = 11$, imply that there is a large difference in the longitudinal and transverse shape of hadron and electron showers. This is useful in particle identification using the calorimeter.

The depth of the EM sections of the ZEUS calorimeter, which will be described in the next section, equals 25 X_0. As mentioned, this is sufficient to contain 98% of the shower energy of a 30 GeV electron. The average energy deposited by hadron showers in such an EM section is approximately 30% of the energy of a 10 GeV hadron, and approximately 20% of the energy of a 100 GeV hadron. Many hadrons do not have their first interaction until after traversing the EM section.

We summarize the showering properties of various particles in a calorimeter with uranium as absorber material.

Muons generally behave like minimum ionizing particles and deposit only a small amount of energy. When they are energetic enough, they will penetrate the calorimeter, leaving only a track-like pattern.

Electrons and other electromagnetically showering particles are fully contained in the calorimeter in a small region. Their showers are narrow and short.

Hadrons produce showers with large fluctuations in structure and size. Before having their first interaction with an uranium nucleus, hadrons may penetrate a

substantial part of the depth of a calorimeter. Their showers are wider and longer than those of electrons. Since every calorimeter has a limited depth, this implies that in showers of highly energetic particles, energy can leak out of the back of the calorimeter. Hadron showers can imitate electron showers, for example through the charge exchange interaction $\pi^+ n \rightarrow \pi^0 p$. The probability for this to happen increases at decreasing hadron energy.

4.3.3 Electromagnetic and Hadronic Energy Resolution

The energy resolution $\sigma(E)/E$ for calorimeters is usually parametrized as:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{\sigma_{noise}}{E} \tag{4.14}$$

where $\oplus$ stands for quadratic addition. σ_{noise} is the contribution of the calorimeter noise to the energy resolution, and is usually small. It is clear that b is the dominating term at high energies.

For EM showers, sampling fluctuations dominate in a, whereas b is determined by the calibration uncertainty and detector imperfections.

For hadronic showers, the contribution to a from sampling fluctuations is small compared to the *intrinsic resolution*, which is related to the energy fluctuations in the nuclear processes that play a role in the shower development in the absorber [40]. These intrinsic fluctuations depend on the type of absorber material.

An additional contribution to b is related to a possible unequal response of the calorimeter to electrons and hadrons. Each hadronic shower has an EM component. The fraction of energy in this component increases with shower energy, but fluctuates in a non-Gaussian way. When the calorimeter does not respond equally to this EM component as to the purely hadronic component, the energy measurement of hadronic showers is severely affected by these fluctuations, which contribute to the factor b in the energy resolution. It is clearly desired to minimize this effect, by making the calorimeter response to electrons equal to the response to hadrons ($e/h \rightarrow 1$), a process known as *compensation*. This can be achieved in hardware by a suitable choice of absorber and scintillator material, and a suitable choice of their relative thickness [2].

The e/mip ratio can be lowered by cladding of the heavy absorber with a small layer of lighter material. This will prevent low energy photons created in the heavy material (but to a lesser extent in the light material) from reaching the active layers. The calorimeter response to hadrons is very sensitive to the production and detection of neutrons in the shower. Neutrons are produced in great abundance in hadronic showers in heavy materials like lead or uranium. They only loose energy by strong interactions with nuclei, by which the total number of neutrons

[2] When the shower development in depth is sampled frequently, the individual electromagnetic and hadronic contributions can in principle be unfolded. In this way, for a non-compensating calorimeter, an equal response to electrons and hadrons can also be reached by software methods.

is greatly increased. The last step in the life of a neutron is capture, but this only occurs when the neutron is practically thermalized. This thermalization is mainly achieved by elastic scattering. When scattering off a uranium nucleus, the energy loss is very small due to the large mass difference, but when scattering off free protons, the energy is quite effectively transferred. Therefore, the energy carried by the neutrons is effectively detected when there are many protons in the active material, as is the case in scintillator. Hadrons and electrons differ in their energy dissipation: a part of the hadronic energy is wasted in nuclei and cannot be recovered. On the other hand, mainly through the neutrons, hadronic showers dissipate relatively more ionization energy in the scintillator. This can lead to compensation. The effect is very sensitive to the ratio of thicknesses of absorber and scintillator plates. For the ZEUS calorimeter, this ratio has been chosen such as to obtain the best possible energy resolution. As an additional effect, the amount of energy carried by the neutrons in a hadronic shower is strongly correlated with the amount of energy lost in binding energy. Effectively measuring this energy reduces the intrinsic fluctuations that contribute to the energy resolution.

4.4 The ZEUS High Resolution Calorimeter

The ZEUS collaboration has chosen for a compensating sampling calorimeter with uranium as an absorber and plastic scintillator as an active material [32, 35, 41, 42]. The basic sampling unit consists of a 3.3 mm uranium plate wrapped in 0.2 mm (EMC) or 0.4 mm (HAC) stainless steel, and a 2.6 mm thick scintillator plate wrapped in paper. This ratio of absorber/scintillator has been chosen in order to obtain compensation. The sampling fraction for electrons is 3.7 %, with an $e/mip = 0.62$. The uranium plates are actually made out of depleted uranium (DU), which in the ZEUS calorimeter is an alloy of 98.4% ^{238}U, 1.4% Nb (which makes the alloy harder), and $\leq$ 0.2% ^{235}U. The scintillator material used is SCSN-38. The produced light is read out via wavelength shifter bars (2 mm thick), and guided to photomultiplier tubes that convert the light to an electrical pulse, that is processed by the front end electronics.

4.4.1 Mechanical Structure

The layout of the calorimeter is shown in figure 4.7. It actually consists of three parts:

- The forward calorimeter (FCAL), $2.2° < \theta < 39.9°$
- The barrel calorimeter (BCAL), $36.7° < \theta < 129.1°$
- The rear calorimeter (RCAL), $128.1° < \theta < 176.5°$

where θ denotes the polar angle with respect to the proton beam axis. The calorimeter covers 99.8 % of solid angle in the forward direction, and 99.5 % in the

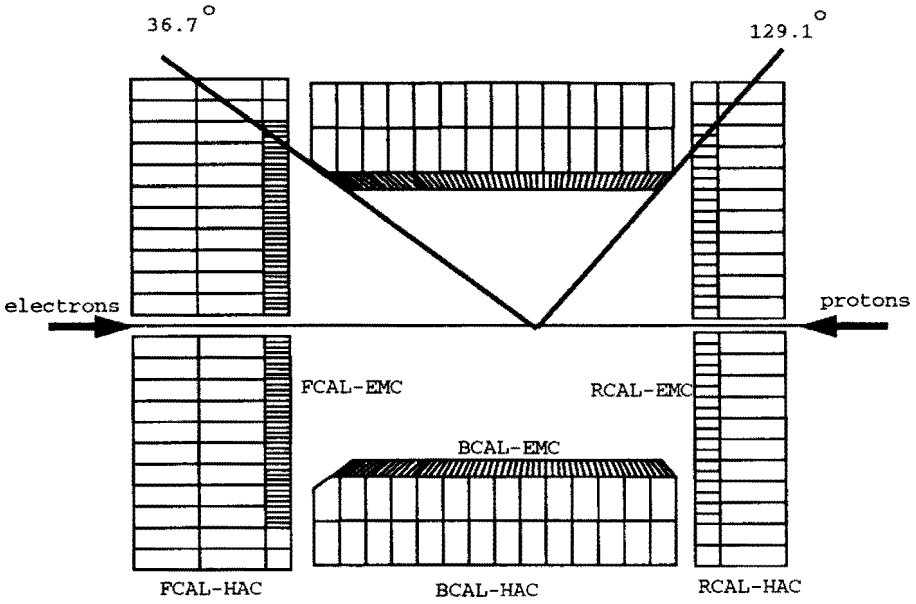

Figure 4.7: *Layout of the uranium calorimeter*

backward direction. In figure 4.8 the angular regions covered by FCAL, BCAL and RCAL are drawn superimposed on the x, Q^2 plane. The left part shows the regions covered in the electron scattering angle θ, whereas the right part shows the regions covered in the quark scattering angle γ. The hatched areas correspond to regions not covered by the calorimeter (beam hole).

The calorimeter will now be described in global ZEUS coordinates, which coincide with the HERA coordinate system as defined in figure 2.12. The layout of the ZEUS calorimeter reflects the HERA beam momentum asymmetry, by the different positions of FCAL and RCAL with respect to the interaction point, and their different depths. The BCAL consists of 32 wedge-shaped modules, placed symmetrically around the beam axis, as shown in figure 4.9. In this configuration, each module covers 11.25° in azimuth ϕ. The modules have a length of 332 cm, and a width of 23.4 cm at the inside, and 43.5 cm at the outside of the active volume. The inner radius of the modules is 122 cm, the outer radius of the active volume equals 229 cm. The modules do not point exactly to the beam axis, but are tilted by 2.5°, in order to prevent cracks between modules pointing to the interaction vertex. The uranium/scintillator layers extend in the $z\phi$ plane, the modules are read out with wavelength shifter bars at the high-ϕ and low-ϕ sides of the modules.

Of each module, the inner 21 uranium/scintillator layers are read out separately, and serve as electromagnetic calorimeter (BEMC). Including the front plate, the

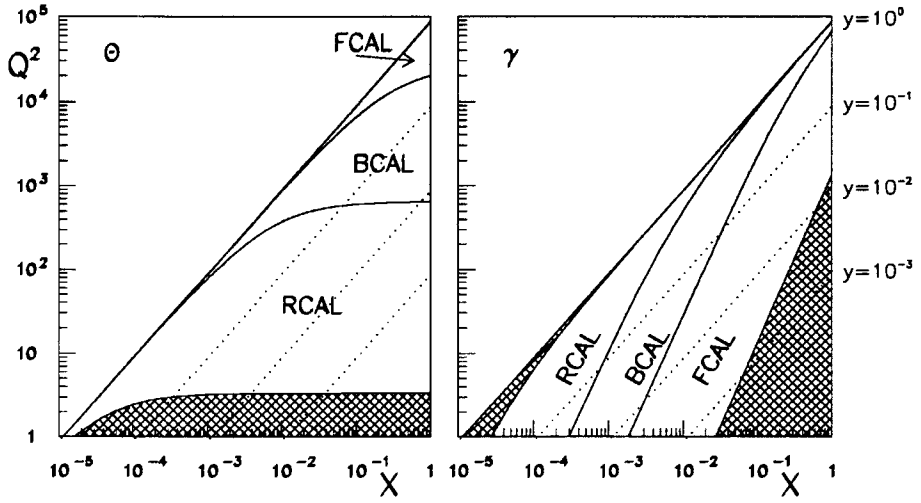

Figure 4.8: *Regions in electron scattering angle θ and in quark scattering angle γ covered by FCAL, BCAL and RCAL, superimposed on the x, Q^2 plane. The hatched areas correspond to the beam holes.*

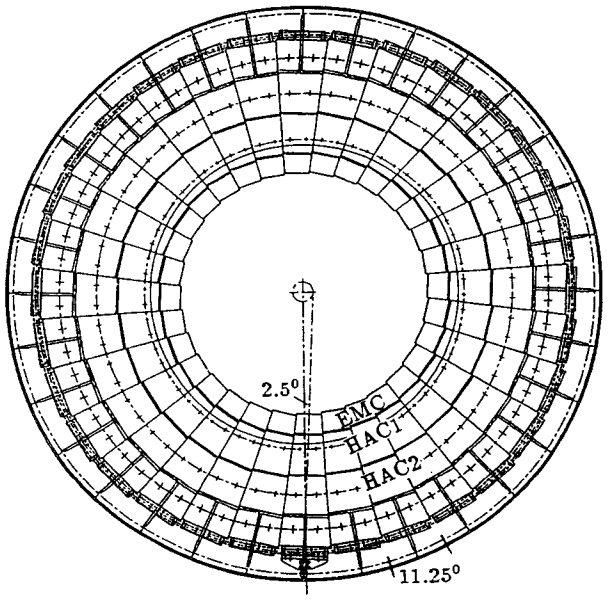

Figure 4.9: *The BCAL seen in the xy plane*

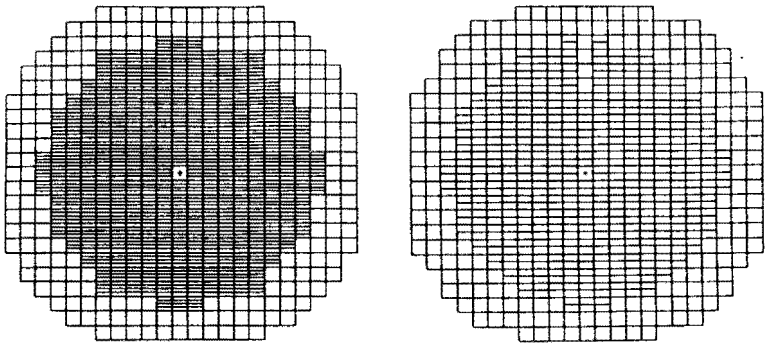

Figure 4.10: *FCAL and RCAL fronts as seen from the interaction point.*

depth of the BEMC corresponds to 22 X_0, or 1.1 λ. The outer 2×49 layers form the hadronic sections BHAC1 and BHAC2, of 53 X_0 or 2.0 λ each. Of each module, the BEMC is subdivided in the z-direction into 53 *cells*, each of which is pointing in θ to the nominal interaction point. Each cell has the shape of a polygon, with dimensions 5.06 by 23.4 cm^2. Each HAC section is subdivided in z into 14 non-pointing cells. Cells located behind each other in depth r form a *tower*. A tower consists of one HAC2 cell, one HAC1 cell, and generally 4 EMC cells. The first BCAL tower, at the RCAL side, only contains 3 BEMC cells, the last tower, at the FCAL side, only 2.

The FCAL and RCAL are mechanically very similar to each other. Each consists of 23 modules, placed vertically next to each other, forming a front facing the nominal interaction point. Figure 4.10 shows FCAL and RCAL fronts as seen from the interaction point. The modules have a width of only 20 cm, and a length varying between 220 and 460 cm. The longest modules are placed close to the beam pipe (with two half modules above and below the beam pipe), the shorter modules at the outer position, so that the front face approximates a circle such as to overlap with the BCAL. In this configuration, the uranium/scintillator layers extend in the xy plane, with the wavelength shifter bars mounted at the high-x and low-x sides of the modules. The first 25 layers, closest to the interaction point, form the electromagnetic calorimeter (FEMC and REMC). This calorimeter has a depth of 25.9 X_0, or 0.96 λ [3]. The remaining layers form the hadronic calorimeter, which consists of 1 unit (RHAC1) in the case of RCAL, and 2 units (FHAC1 and FHAC2)

[3] The geometry of a sampling unit slightly differs between FCAL/RCAL and BCAL. This change of geometry has a different effect on the value of X_0 as it has on the value of λ.

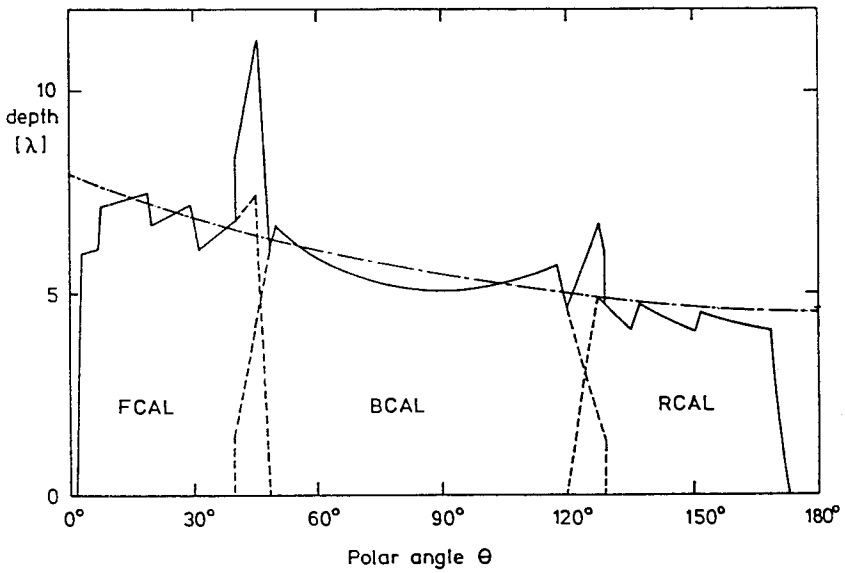

Figure 4.11: *Depth of the calorimeter in λ as a function of polar angle θ.*

in the case of FCAL. Most of these sections have 80 layers, corresponding to 3.09 λ. Some of the outer modules are less deep: some RHAC1 sections have only 60 layers (2.32 λ), some FHAC2 sections have 60 or even 40 layers (1.54 λ). Each module is segmented in y in cells. HAC cells measure 20 by 20 cm^2, FEMC cells 5 by 20 cm^2, and REMC cells 10 by 20 cm^2. Again, cells behind each other form a tower, of 1 HAC1 cell, 1 HAC2 cell (only FCAL), and 4 EMC cells (2 in case of RCAL). The outer parts of the FCAL and RCAL modules do not have the subdivision of the first 25 layers in 4 or 2 cells, since they are shielded by the BCAL and do not have to serve as electromagnetic calorimeter; these cells are called HAC0. A part of the R1T module (RCAL above the beam pipe) lacks EMC cells in order to make place for feed pipes for the cryostat, this part is known as the chimney.

Each calorimeter module is equipped with wavelength shifter bars on both sides. In this way, each individual calorimeter cell is read out by two photomultiplier tubes, or PMT's, on different sides.

The total number of channels (PMT's) in the calorimeter is 11836. Since each cell is read out by 2 PMT's, the total number of cells equals 5918. These can be subdivided in 1056 FEMC, 196 FHAC0, 460 FHAC1, 460 FHAC2, 1696 BEMC, 448 BHAC1, 448 BHAC2, 512 REMC, 190 RHAC0 and 452 RHAC1 cells.

Figure 4.11 shows the depth of the calorimeter in interaction lengths λ as a function of polar angle θ. The depth is sufficient to contain 95% of the energy for 90% of the jets with maximum energy allowed by HERA kinematics, as indicated by the dash-dotted line [39].

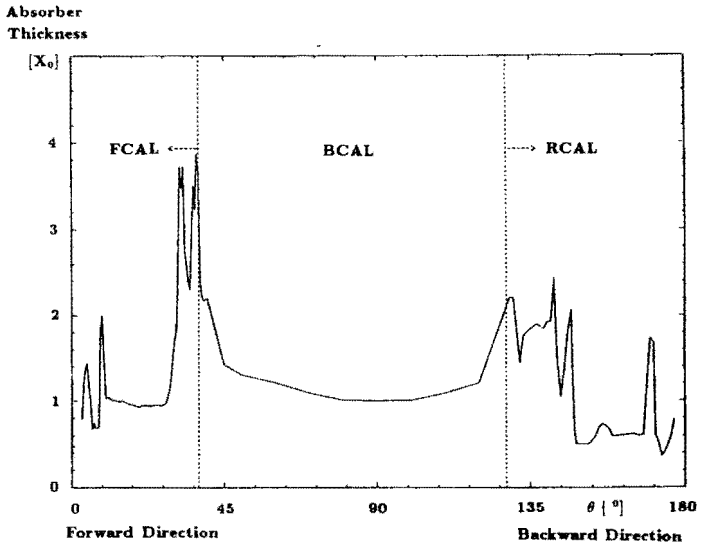

Figure 4.12: *Amount of dead material, in X_0, in front of the calorimeter, as a function of polar angle θ.*

The inactive, or *dead*, material between interaction point and calorimeter (detector flanges, support structures, cables) has a negative influence on the response of the calorimeter. Figure 4.12 shows the amount of dead material, in units of radiation lengths X_0, in front of the calorimeter, as a function of polar angle θ [39]. For most angles, the amount of dead material does not exceed 1-1.5 X_0, but at some places, this can be as much as 4 X_0.

The Hadron Electron Separator

The hadron–electron separator, or **HES**, is a subdetector of ZEUS mounted inside the uranium calorimeter. It consists of several single layers of silicon diodes, located after 3.3 (FCAL and RCAL), and 6.3 (FCAL only) radiation lengths in a gap in the calorimeter. This location is chosen such that the HES diodes are located close to shower maximum of electromagnetic showers. At this location, showers of particles like electrons and photons consist of many particles, and thus give a large signal in one, or a few, HES diode(s). Hadrons, in contrast, behave like mips in the HES, since they usually have their first interaction vertex at greater depths. The diodes have a dimension of 3 $\times$ 3 cm^2. This granularity is better than the calorimeter segmentation, and considerably increases the position resolution of showers, and the two track resolution. This is of special importance for the electron measurement in the rear direction.

4.4.2 Readout of the ZEUS calorimeter

The readout of the ZEUS calorimeter has to fullfil the following design criteria [43]:

- The dynamic range is determined, at the low end, by the expected noise levels that are used to monitor the stability of the calorimeter, which equal approximately 10 MeV, and, at the high end, by the maximum energy to be recorded in a single channel, which can be as high as 400 GeV in FCAL.
- Since hadronic showers develop relatively slow compared to electron showers, and there is a considerable amount of energy in the shower tail, the measured energy for hadrons depends on the gate time. The effective gate time needed to reach compensation is roughly equal to the time between HERA bunch crossings of 96 ns [4]. The readout must be able to handle this speed.
- The first level trigger needs a time of a few μs to reach a decision, which is much longer than the time between bunch crossings. The calorimeter signals must in the mean time be stored in a pipeline.
- The time of arrival of the pulses must be measured with a resolution of $\mathcal{O}(1)$ ns, in order to distinguish between ep events and upstream beam-gas interactions (see section 8.5).
- The first level calorimeter trigger must have fast access to the signals.
- For calibration purposes, the DU current must be measured by integrating the signals over a time of 20 ms (see section 4.4.3).

12 Photomultiplier tubes, located at one single side of a module, are connected to an analog front end card mounted on the modules. Digital cards, located in the Rucksack, digitize the signals after the event has been accepted by the first level trigger. At normal data taking, energy and time are calculated on-line (i.e. real-time) for each channel, by dedicated processors (DSP's) on the digital cards. After acceptance by the second level trigger, readout processes running on transputers pass the event data to the event builder.

Analog Cards

The layout of an analog front end card is shown in figure 4.13. The signal of a PMT is split into 4 circuits. One (RT) is used for triggering and leads to the calorimeter first level trigger electronics. A second one is used for integrating the DU current for calibration purposes. The remaining two circuits are used in the actual readout.

The dynamic range needed in order to reach the first design criterion is larger than what can be obtained with a 12 bits ADC. In order to solve this, the signal of

[4] In fact, compensation depends on both the gate time and the ratio of absorber/scintillator thicknesses. For ZEUS, the gate time was set by the time between HERA bunch crossings, and the thickness ratio was optimized keeping this in mind.

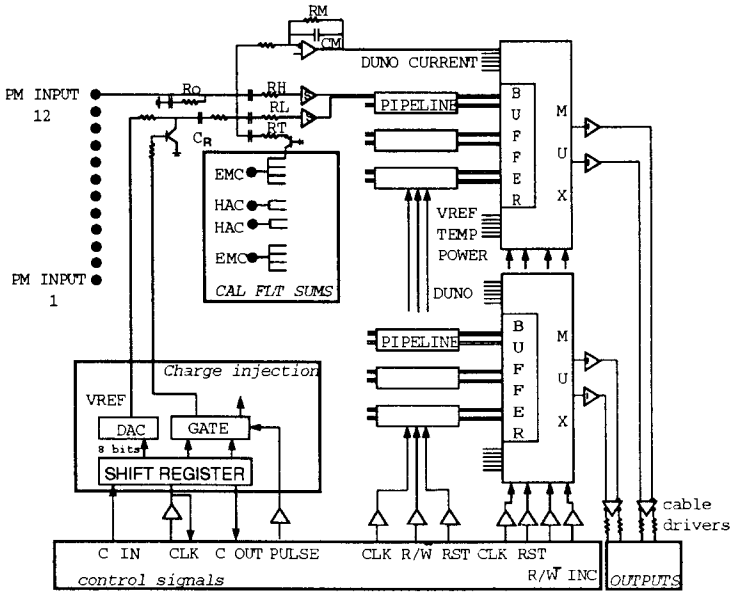

Figure 4.13: *Layout of an analog front end card.*

each PMT is split into a low gain (RL) and a high gain (RH) channel, which each cover part of the total energy range: the high gain channel the lower energies up to $\mathcal{O}(20)$ GeV, and the low gain channel the higher energies. The original PMT signal pulse, as shown on the left of figure 4.14, is shaped by a pulse shaper circuit (S) into a pulse as shown on the right of figure 4.14 [43]. The result is a stretched pulse with fast restoring of the baseline for the acceptance of new events [32]. The signal is subsequently stored in a 58 cell switched capacitor analogue pipeline (for the need of this pipeline, see section 4.6); the pipeline accepts a sample of the signal every 96 ns. For the pulse shown in figure 4.14, the sampling times are denoted as t_i; we shall write the function describing the pulse shape as $h(t)$, and the resulting stored samples as h_i.

When an event is accepted by the first level trigger, the pipeline clock is stopped, and up to 8 samples, corresponding to the accepted event, are transferred to a buffer. The samples of 6 PMT's are multiplexed and sent to the a digital card. The samples will be used there to reconstruct energy and time; at least 3 samples are needed to do so, and additional samples can be used to detect any pile-up effect.

This procedure is performed for every channel in high gain as well as in low gain mode, samples for both modes are sent to the digital cards.

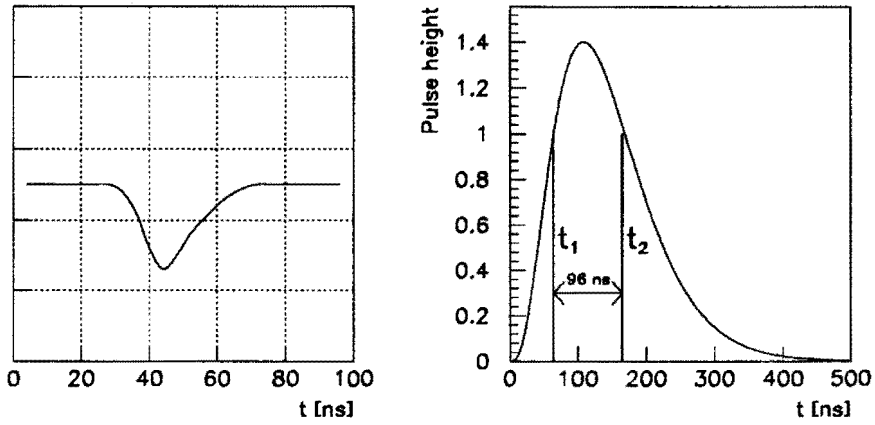

Figure 4.14: *The original PMT signal pulse (left), and the pulse after shaping (right). The pipeline samples the pulse every 96 ns, at times t_i.*

Digital Cards

The main components on a digital card are the ADC's, and a dedicated digital signal processor (DSP). 4 ADC's are placed on a digital card, and each ADC digitizes the samples of 6 PMT's. The DSP is able to reconstruct energy and time on-line for each channel from the digitized samples. Actually, the DSP reconstructs the integrated charge of the PMT pulse, but this is proportional to the deposited energy. For most channels, the samples taken in high gain are used for the reconstruction; when an overflow occurs in the calculation, the low gain samples are used. After calculation of energy and time, the DSP makes this information available to the second level trigger processors and the readout processes.

At least three samples h_i are needed to reconstruct energy and time. One of the samples, which we call h_0, is situated on the baseline. Two other samples, h_1, taken at time t_1, and h_2, taken at time t_2, are situated on the rising and falling edges of the pulse, as shown on the right in figure 4.14. The DSP energy reconstruction method tries to reconstruct the deposited charge of the pulse under the approximation that the shaper pulse has a triangular shape in a short time range [32]. The reconstructed charge is:

$$Q = (h_1 - h_0) + C_R(h_2 - h_0) = H_1 + C_R H_2 \tag{4.15}$$

where C_R is defined as

$$C_R = \frac{\frac{dh}{dt}\Big|_{t=t1}}{-\frac{dh}{dt}\Big|_{t=t2}} \tag{4.16}$$

Thus, C_R is a constant, and equal to 1.80. The time can be reconstructed as:

$$T = \frac{H_1 - H_2}{\frac{dh}{dt}\Big|_{t=t1} - \frac{dh}{dt}\Big|_{t=t2}} = \frac{H_1 - H_2}{C_D Q} \tag{4.17}$$

with C_D defined as:

$$C_D = \frac{\frac{dh}{dt}\Big|_{t=t1} - \frac{dh}{dt}\Big|_{t=t2}}{Q} \tag{4.18}$$

Both reconstructed charge and time depend on the timing of the pulse relative to the sampling clock, therefore the above equations only hold if the pulse is exact in time, i.e. $h_1 = h_2$. It is possible to correct charge and time with a polynomial, corresponding to a Taylor expansion around nominal time [44]:

$$Q_{corr} = Q(1 + \sum_{i=1}^{4} a_i T^i) \tag{4.19}$$

$$T_{corr} = \sum_{i=1}^{3} b_i T^i - T_{offset} \tag{4.20}$$

where T_{offset} is a time offset for each individual channel that needs to be taken into account; this offset is related to effects such as different wavelength shifter and light guides lengths and different transit times in PMT's and electronics. In addition, the geometrical location of the cell (i.e. the time of flight of a particle between vertex and calorimeter) plays a role. The offsets are chosen such that all channels with energy from an actual ep interaction in the nominal interaction point give $T = 0$. The factors a_i and b_i are constants that are determined in a test beam.

With the time resolution that can be achieved in this way, the calorimeter is able to assign tracks to a specific beam crossing. Neither the inner ZEUS detectors nor the backing calorimeter or the muon chambers reach this resolution. Therefore, the calorimeter sets the time of an event, also for the other detectors.

Transputer Readout

The second level trigger and readout processes are running on a transputer network, which has the advantage of a high connectivity and parallellism [45, 46]. A cluster algorithm is running on the second level trigger processors, and its results are sent to the global second level trigger. After obtaining a positive second level trigger decision, the readout transputers gather the data from the digital cards, format the data into ZEBRA banks, and send it to the event builder, where it is ADAMOized.

4.4.3 Calibration of the ZEUS Calorimeter

The constraints on the accuracy of kinematics reconstruction in DIS events demand the energy scale of the calorimeter to be known within a few percent at any time. Therefore, the calibration of the calorimeter response is of crucial importance to ZEUS.

An overview of the calibration methods for the ZEUS calorimeter is given in figure 4.15.

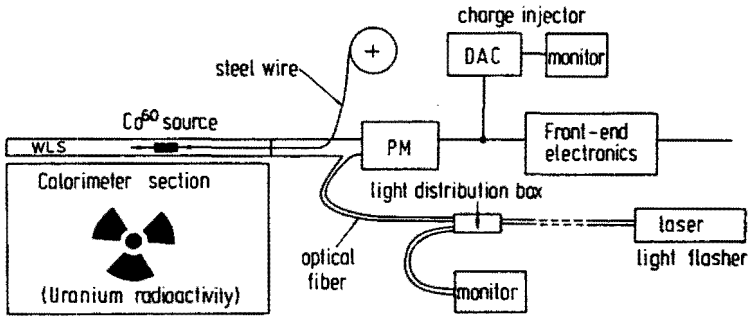

Figure 4.15: *The various calibration tools for the ZEUS calorimeter.*

As described in the previous section, the DSP's calculate, at normal data taking, energy and time for each channel on-line. In order to be able to give this output in units of MeV c.q. ns, the DSP's need a vast amount of calibration constants related to the front-end electronics, and the value of C_D and C_R of equations 4.18 and 4.16 and the parameters of the polynomials of equations 4.19 and 4.20 [47]. Since the constants for the front end cards include pedestal voltages and gain factors for every single pipeline cell and buffer cell, in high gain as well as in low gain mode, the total number of constants exceeds 275 per channel. These electronics related calibration constants are not really constant, but vary slowly over time. They need to be calculated from regularly taken calibration data. For this electronics calibration, a programmable charge injection system injects a known amount of charge to the front-end electronics, and a programmable DC voltage can be applied to the inputs to the pipelines. In addition, the validity of the pedestals is monitored by taking empty (random) triggers. By taking regular calibration runs, and monitoring the constants and their variation, stability can be maintained.

Before installation in the ZEUS detector, each FCAL and RCAL module has been scanned with a movable radioactive Co^{60} source. Such a source can detect aging effects in the wavelength shifter, as well as defects in the stacking.

An important advantage of uranium calorimeters is the possibility to use the uranium natural radioactivity (uranium noise, or UNO), as a calibration tool. For this purpose, the uranium induced PMT current is integrated over a relatively long time (20 ms in order to create a stable signal) using integrators on the analog cards. The magnitude of this current depends on the size of the calorimeter cells, the physical uniformity of these cells and the optical elements, and the high voltage level applied to the PMT's. Great care has been taken in the assembly of the calorimeter to control the quality of the calorimeter elements; remaining factors that may affect the calorimeter signal, such as non-uniformities and defects in the cells and the optical elements, cancel in the ratio $\frac{particle signal}{UNO}$. This makes the UNO an excellent calibration tool. The energy scale of all channels is set by tuning

the individual high voltage levels in such a way that the uranium currents for all equally sized cells is equal. For cells of unequal size, such as EMC and HAC cells, the UNO settings differ.

Due to the long lifetime of uranium nuclei ($4.5 \cdot 10^9$ y), and the large amount of uranium nuclei in the calorimeter, the UNO provides an extremely stable signal. For monitoring purposes, the UNO currents are remeasured on a regular basis. Since the value of the UNO current is directly related to the PMT gain, and thus the energy scale, it needs to be monitored carefully. The short term variations (few hours) in the UNO have been shown in tests to be small: typically between 0.1 and 0.5% [48]. Therefore, one may conclude that the short term PMT gain variations, and therefore the uncertainties in the energy scale, are small. The long term variations (many days) are larger, in the order of 3% [48], but by monitoring them, their effects on the data that has been taken with the calorimeter can trivially be corrected offline (UNO calibration).

During normal data taking, the uranium radioactivity signal acts as a background (noise) to the particle signal. This will be investigated in section 6.5.

A subset of FCAL, BCAL and RCAL modules have been calibrated in electron, hadron and muon beams at CERN and FNAL [41, 42]. From this data, the absolute energy scale, i.e. the conversion factor between reconstructed pulse heights (in pC) and corresponding particle energy (in GeV) has been obtained. In addition, all FCAL and RCAL modules have been scanned in a cosmic ray muon stand, and cosmics and beam halo muons are continuously used in *in situ* calibrations at HERA.

The module to module variations have been shown to be in general less than 1% [32], so that there is no problem in transferring the energy scale to other, not calibrated, modules. After full calibration, the energy resolution for electrons in these tests have been measured to be $\sigma(E)/E = 0.18/\sqrt{E}$ [32]. These single modules, however, do not fully contain hadron showers, so that the energy resolution for hadrons is measured to be $0.35/\sqrt{E}$ from the data taken with a prototype calorimeter [48]. The constant term contributing to σ/E has been shown to be less than 0.02 for the prototype, and 0.0035 for electrons in the real modules.

The calorimeter is equipped with a light flasher system. This system is used for testing (and correcting) the linearity of photomultipliers and readout electronics, since the system is able to operate in a large dynamic range. Light from a laser or LED is distributed via optical fibers to the modules, and inserted into the light guides just in front of the PMT's. Monitoring diodes measure the intensity of each flash. Since the timing of the light pulses is known, the light flasher system plays an important role in the calibration of the time measurement. Individual time offsets for each channel can be measured.

The linearity of the PMT response has been measured at CERN in high energy electron beams. For the FCAL, no deviation from linearity was observed, whereas for RCAL, with a different type of PMT's, a deviation of 2% was observed at 100 GeV [39].

The longitudinal and lateral shower containment was measured in test beams, and agrees with theoretical expectations.

Test beams have also been used to check the spatial uniformity of the calorimeter response. In particular three sources of nonuniformities have been identified [48]:

- The boundary between modules. At normal incidence, electrons entering the calorimeter via the wavelength shifter material between the modules, have a considerably higher response than those entering the center of a module. These can be reproduced by Monte Carlo, assuming a certain contribution of the light created in the WLS to the total signal. Hadrons do not show such an effect, and the effect is also reduced at incidence under non-zero angle. In ZEUS, these cracks can only be reached at angles of 40, 120, 200 ... mrad. It has been shown that the inclusion of lead foils between the modules reduces the effect to an acceptable level of less than 5% at normal incidence.

- The location of the spacers. Spacers are located at the corners of each tower. At such places, the electron response is reduced by about 20%. The spacers, however, only occupy a very small part of the active area of the calorimeter, and in ZEUS, electrons always enter the calorimeter under a non-zero angle.

- The boundary between EMC sections in a given module. A gap of 1 mm between scintillator plates is responsible for a 5% decrease of the calorimeter response. In ZEUS, these gaps are 0.6 mm, and lead to a 3% decrease at normal incidence, which, again, is not a common situation in ZEUS.

4.5 Particle Identification in ZEUS

The main requirements for particle identification in HERA physics lie in the recognition of electrons and muons against the large background of hadrons.

Muons are best recognized by their large penetration power since they behave approximately as mips. Provided their momentum exceeds about 2 GeV, they will penetrate through the calorimeter and be measured in the muon chambers. Therefore, they can be recognized as a track in the muon chambers, matching with a track in the inner detector and a minimum ionizing energy deposit in the calorimeters. Many of the muons produced in collissions at HERA, however, have an energy of less than 2 GeV, and are stopped in the calorimeter.

For the recognition of electrons, the ZEUS detector offers a number of possibilities:

- The shape of the shower in the calorimeter. Due to the large difference between the scale of an electromagnetic shower (X_0) and a hadronic shower (λ_{int}), there is a striking difference between showers of electrons and hadrons in the calorimeter. This, and its use in particle identification, will be further discussed in section 6.11.

- The measurement of dE/dx in the tracking detectors. The dependence of the energy loss dE/dx of a charged particle on its velocity can be used for particle identification up to energies of $\mathcal{O}(10)$ GeV.
- The signals of the transition radiation detector.
- The pulse heights in the Hadron-Electron Separator diodes. The fact that the HES diodes are located close to shower maximum of electromagnetic showers in the calorimeter, implies that electrons and photons induce large signals in the HES, whereas hadrons behave, for the HES, like mips.

4.6 Trigger and Data Acquisition

The requirements of a trigger for HERA result from the following points:

- The background event rates that can be expected from the machine and from cosmics lie between 10 and 100 kHz;
- The rate at which events can be written to tape lies between 5 and 10 Hz. This requires a reduction of a factor between 1000 and 20000. The accepted events will be dominated by photoproduction events at low or medium p_T, which are interesting by themselves, but also form a large background to even more interesting events;
- The expected size of raw events lies between 100 and 200 kByte;
- The time between bunch crossings is 96 ns. This is too short to reach a first level trigger decision. Therefore, the data of various bunch crossings needs to be stored in a pipeline. A reasonable time for a first level trigger decision is 5 μs, which means that the pipeline must be some 50 to 60 events deep. Also the trigger itself must be designed to operate on pipelined trigger information, because the time needed in the trigger to process the data exceeds the trigger rate.

The beam induced backgrounds have been described in section 3.3.

The ZEUS trigger system is designed to have three stages [34, 35]. A schematic layout of these stages is shown in figure 4.16. On the first trigger level, the data is analogue. The second level trigger is the first level where digitized data is available. Between second and third level trigger, an *event builder* collects all data from the components. The third level trigger consists of a CPU-farm running offline-like algorithms on the full data.

The task of the first level trigger (FLT), is to reduce the trigger rate to $\mathcal{O}(1)$ kHz. The (analogue) data is first processed locally by the component first level trigger systems, and they subsequently send their information to the Global First Level Trigger (GFLT), which is equipped with programmable logic to combine this

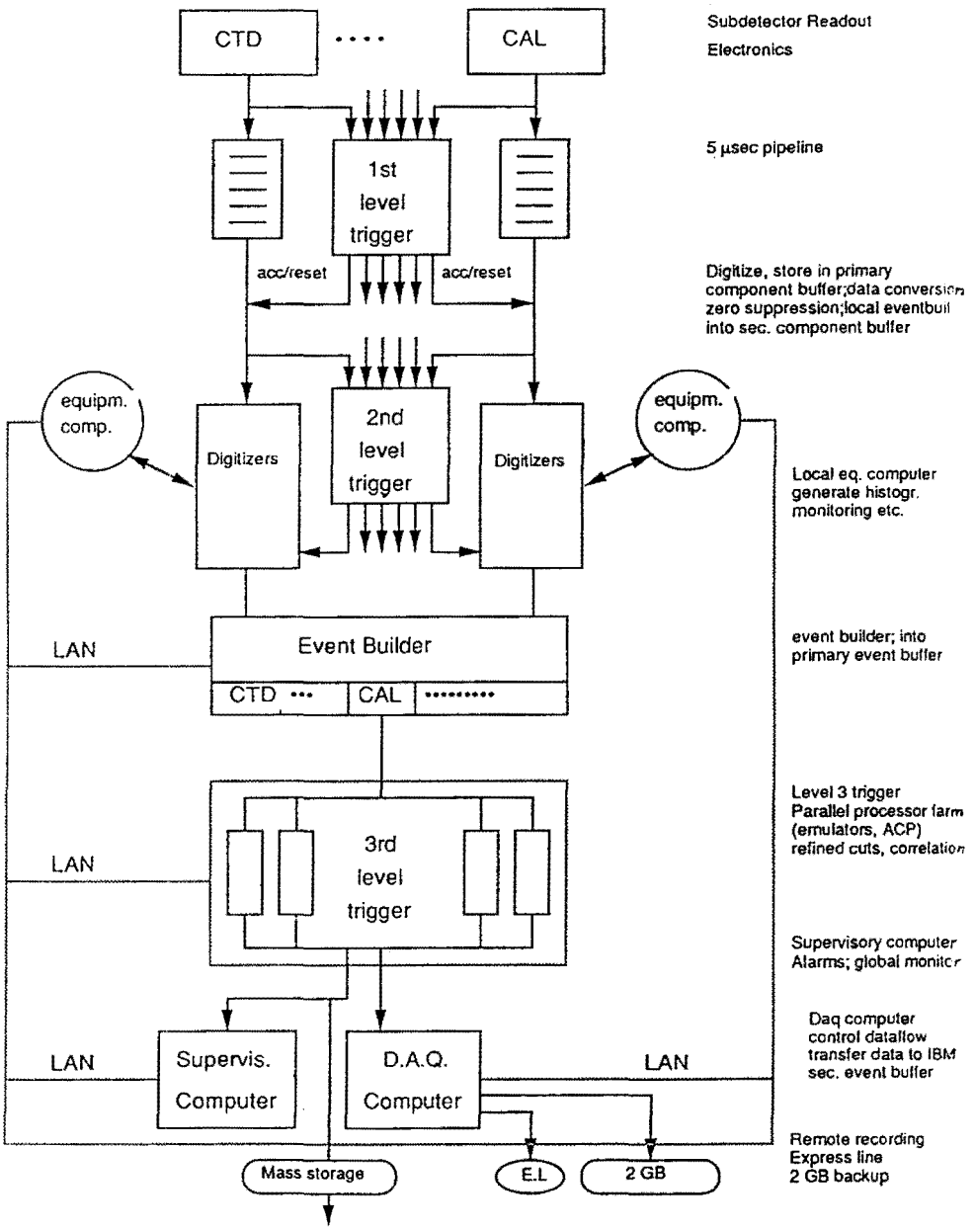

Figure 4.16: *Layout of the ZEUS trigger system.*

information in a very large number of ways. This enables the setup of a very flexible system.

The calorimeter FLT works with data coming from the front end electronics. By appropriate summation, information on both global and local energy deposits can be obtained, which includes total energy, transverse energy and transverse momentum, EMC and HAC energies, energies near and away from the beam pipe, energy clusters, quiet flags and candidates for isolated electrons and muons in 16 calorimeter regions.

After a first level trigger and after the data has been sent to the ADC cards, the data is digitized and ready to be processed by a second level trigger processor.

On the second trigger level, computer algorithms can run on fast processors operating on the digitized data. The SLT is designed to reduce the trigger rate to below 100 Hz.

The first few ms are again in principle available for component second level trigger processors. Of the components, the calorimeter and the inner tracking detectors run complicated clustering and track finding algorithms on their component SLT's.

The resulting clusters and tracks can, in combination with data from other components, be used by the Global Second Level Trigger (GSLT) to calculate a trigger decision. On this level, also timing information from the CAL is available.

A very large part of the SLT hardware is based on *transputers*, allowing a high degree of connectivity and parallellism.

When a positive GSLT decision is taken, the components read out the digitized data from the digital cards, and send it to the event builder, which collects the data from all components into one event. The components format their data as ZEBRA banks, for further processing this format needs to be adapted slightly in order to be compatible with the file format used by ADAMO (see next chapter). A complete event is passed to the third level trigger farm.

The third level trigger (TLT) has to reduce the trigger rate to a level acceptable for data storage and offline processing. The TLT consists of a farm of Silicon Graphics RISC computers running in parallel. A subset of the offline reconstruction program, in addition to specially written fast track and vertex finding routines, operates on the full event data. Since a complete offline environment is available, the system is very flexible and new algorithms, for example for unforeseen effects (sparks), can be tested and implemented on a very short time scale.

After an event has been accepted by the third level trigger, it is stored for further offline reconstruction. For this purpose, the third level trigger farm is connected to the main DESY IBM on the DESY site via an optical link, over which data can be written to tape at a rate of $\mathcal{O}(1)$ MByte/s.

Chapter 5

ZEUS Offline Software

5.1 The Offline Software Chain

This chapter is devoted to a description of the ZEUS offline software, and the structure of the offline event reconstruction program. A detailed description of the calorimeter reconstruction software is given in the next chapter. Whereas the ZEUS interface to the Monte Carlo event generator programs will be described here, the details of the generators used in this thesis will be described in chapter 7.

The ZEUS offline software consists of a number of programs. The connection of these programs and the data flow between them is described in figure 5.1.

- ZDIS [49] is a program for generating events on the 4-vector level by providing an interface to Monte Carlo event generators such as LEPTO, PYTHIA, etc.
- MOZART [50] is the ZEUS detector simulation program. The 4-vector output of ZDIS can be fed as an input to MOZART.
- ZGANA simulates the ZEUS trigger logic. As input it accepts simulated data from MOZART as well as real ZEUS event data.
- ZEPHYR is the ZEUS Physics Reconstruction program. Real ZEUS event data, as well as MOZART data, optionally processed by ZGANA, is read as input.
- EAZE forms a frame for convenient physics analysis.
- GAZE [51] and LAZE [52] are event display programs.

The programs have many things in common:

- All programs make use of ADAMO [53] as data manager, whereas the detector description is performed using the tools provided by GEANT [54], and the memory management is handled by ZEBRA [55].

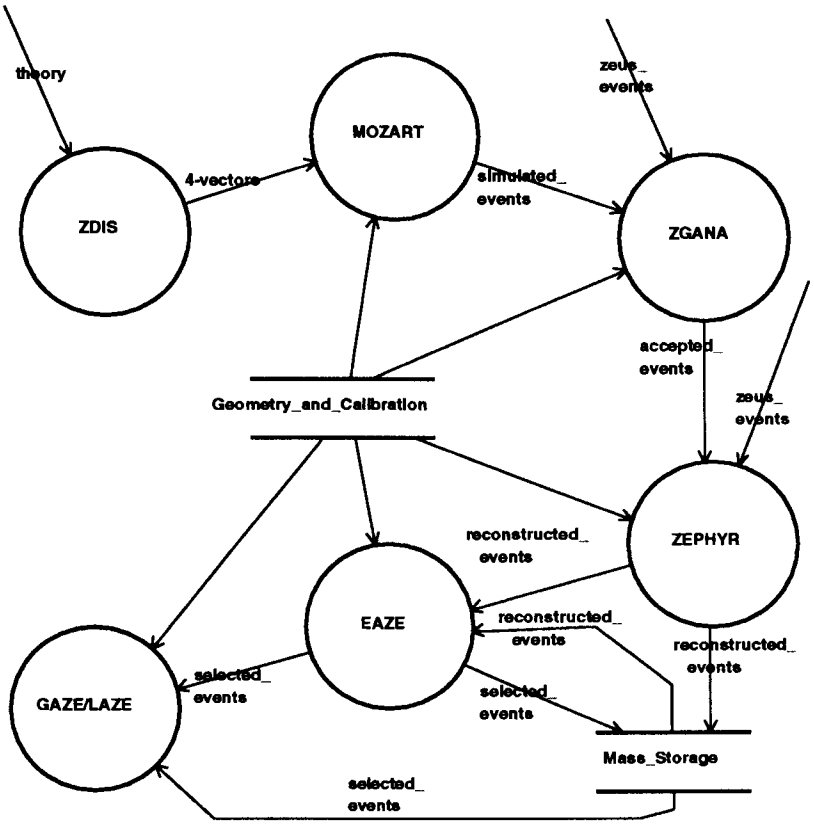

Figure 5.1: *The ZEUS offline programs and the data flow between them.*

- The output data of all programs is in an identical format known as General Adamo Files, or GAF's. For data on files, ADAMO makes use of the ZEBRA FZ or RZ formats in native or exchange modes. This ensures that the data is easy to exchange between various computers.

- All programs make use of the same geometry and calibration constants base, which is also set up using ADAMO [56].

- A number of packages written on top of ADAMO [57, 58], taking care of data card reading, file handling and program monitoring, are commonly used by all programs.

The real event data, as measured by the ZEUS detector, is also formatted as a GAF by the event builder.

In the next sections, we will review the concepts of the software design techniques

of SASD, and the data manager ADAMO. After that, the implementation of these concepts in various packages and interfaces in the reconstruction software will be discussed. Finally, we will describe shortly some of the programs of the ZEUS offline software, and in a bit more detail the reconstruction program.

5.2 SASD

The concepts of Structured Analysis and Structured Design (SASD) comprise a set of methods and rules that can be used in the design, building and documentation of any system, such as a computer program. For a concise summary of these methods and rules, see [59]. SASD is concerned with building a series of models of a system, emphasizing in a well defined and implementation–independent way the vital properties of the system under investigation. By formalizing the way the system specifications are given and the way the system is described, the method tries to make complex systems more transparent, and facilitates the communication in between the designers of the system, and the designers of the system and the outside world. A series of computer aided software engineering (CASE) tools, such as Teamwork [60], implement the SASD methods in system design software packages, with full logical structure checking.

In the design phase, an *essential model* is made, describing the vital properties of a system in an implementation-independent way, by means of a series of submodels. The *environmental (sub)model* describes the environment in which a system is embedded (*context diagram*) and the structure of the data it has to operate on (*entity-relationship diagrams*), whereas the *behavioural (sub)model* deals with the actual operation of the system in terms of data flows and processes acting upon and transforming the data (*data flow diagrams*).

We will limit ourselves in this thesis to context diagrams and data flow diagrams in the description of the software, and to entity relationship diagrams in the description of the data. Figure 5.1 is an example of a data flow diagram. The bubbles represent objects (such as programs) that process and transform data. The data to be processed or produced is represented by arrows, and may be stored in data stores. The structure of that data is described by entity-relationship diagrams, as shown for example in figure 5.2.

The definition of the data structure is of vital importance to offline reconstruction software in high energy physics, which deals with large amounts of data with a complicated structure.

Entity-Relationship Diagrams (ERD's) [61] are concerned with the objects that form the data ('entities'), the properties ('attributes') that make a particular object unique, and the relationships between those objects. One should emphasize that the concept entity applies to the class of similar objects (e.g. particle), not to one manifestation of such an object in particular (e.g. an electron). The attributes are necessary to describe one manifestation of an entity uniquely, i.e. they describe its properties. It is convenient to think of an entity as a table, with the attributes

forming the columns and the various manifestations of the entity forming the rows. The number of columns is always fixed, the number of rows may vary over time. Entities are drawn as boxes in ERD's, with the attributes listed in the box. Relationships are arrows between entities in ERD's, and denote that two or more entities have something to do with each other, and they describe the nature of that relation. Relations can be one-to-one, many-to-one, or many-to-many.

5.3 ADAMO

ADAMO [53] is essentially an implementation of the entity-relationship model for management of data structures in FORTRAN. ADAMO supports the notion of data as a set of interrelated *tables*. ADAMO makes use of ZEBRA for its memory management.

In the program design phase, the structure of the data to be handled by the program needs to be described formally in the form of a DDL (data definition language). A DDL is formally equivalent to an entity-relationship diagram, or a set of such diagrams. It contains exactly the same information on the entities, their attributes and relationships. Therefore, the DDL can be seen as a 1 : 1 image of an ERD, in written language, following a well defined syntax. The DDL can subsequently be transformed, using standard ADAMO tools, in FORTRAN code that sets up the data structure in an application program. Therefore, in the design phase, SASD methods are used to design the program and find the optimum data structure, and through the DDL, this data structure is implemented in FORTRAN by ADAMO.

The application program can now manipulate the data, for example by reading data from a file, changing it, and writing it out to a new file. The data consists of a number of tables, with a fixed number of columns (the attributes), but a variable number of rows (the various manifestations of an entity). Rows can be added to, deleted from, or changed in a table. Indices and selectors provide a very flexible way of accessing rows in table by ordering the table (or a part of a table) along the values of one or more attributes. The access to the data and the most optimal information retrieval is thus guided by the application: different problems require different ways of searching through the data. ADAMO therefore emphasizes the *logical* structure of the data, taking care internally (but not letting the user worry about) the *physical* structure of the data, i.e. where is it stored in memory. Other ADAMO routines enable navigation through the data using the defined relationships: from a given row in a given table, related information in other tables is thus easily retrieved.

As an example it is convenient to think of the structure of calorimeter data. Cells in a calorimeter are characterized by a position, which is fixed, and an energy and time that changes from event to event. The cell is related, through its position, to towers, modules, rotation matrices, digital cards, etc., but also, from event to event, to clusters, jets, reconstructed tracks in a tracking detector, and in a Monte Carlo situation, to generated 4-vectors. The way the information of the cells is accessed

varies: sometimes the position is needed, sometimes the energy, and sometimes the time. The same holds for the related information: sometimes readout information from the digital cards is requested, sometimes the link with a reconstructed jet, and sometimes the energy of a Monte Carlo generated particle that caused the energy deposit. An example of a visualized data structure is shown in figure 5.2. The figure shows the ADAMO tables with their attributes and relationships.

The use of relationships in ADAMO is limited to one-to-one or many-to-one, they form an extra column in the table where the relationship originates. The data structure in memory is in principle for safety reasons not directly accessible: ADAMO handles the actual data structure internally, and gives user access to the logical structure only. Each table can be operated upon row by row. For this purpose the *window common block* is used, which is a standard FORTRAN common block. Each table has one such window common block, that can hold the information of exactly one row of a table, attributes and relationships, and which can be used to communicate with the data structure in memory. The window is the standard interface between the program and the ADAMO system. The program requests the system to display a specific row, acts upon it and may let it be restored. All adding of new rows, and navigation between tables proceeds through the ADAMO windows. As an additional advantage of these window common blocks, each variable has a standard FORTRAN name, and need not be part of some big array.

The input and output of data structures in memory to and from file is handled using General Adamo Files (GAF's), which are based upon the ZEBRA FZ and RZ file formats. The data structure on file is analogue to that in memory, but includes a description of that structure (*metatables*), so that ADAMO can easily handle structures changing over time. The fundamental object on a GAF is the *data flow*, collecting a number of tables. Each data flow is uniquely identified by a *key*. A key is itself an ADAMO table, and collects the information (attributes) of a data flow. In this way, arbitrary numbers of any data flow may appear on a GAF.

The use of ADAMO as a data manager on top of ZEBRA gives additional overhead. It has been shown [62] that this overhead is small, when careful programming is done, since search procedures are optimized. In return, the combined use of SASD and ADAMO makes a program more transparent, easier to code (since more time is spent in the design phase), more flexible, and easier to debug.

5.4 Use of ADAMO in the ZEUS Software

ZEUS uses ADAMO in various ways. The DDL that describes the data structure, as needed by ADAMO, also serves as documentation of ZEUS data. The real ZEUS event data, i.e. the measured detector signals, also follows the GAF format, and therefore the event builder needs knowledge of the DDL. In fact, all data files in ZEUS are in the GAF format, either in FZ sequential (event and Monte Carlo data), or in RZ direct access format (geometry and calibration constants). In the ZEUS offline software, all communication with the data structure in memory goes via the

routines of the TAP (Table Access Package) as formally required by ADAMO.

5.4.1 Event data

The event builder writes out the raw ZEUS event data in ADAMO FZ exchange format. Since all data exchange in ZEUS goes via identically structured GAF's, the data flow between the programs as shown in figure 5.1 is very consistent. The key used for the identification of the data flows on the GAF's holds information like the name of the data flow, the run and event number for sequential data (version numbers for direct access data), and the time of data taking (ranges of validity in time for direct access data). A Monte Carlo generated event, after detector simulation, typically has three data flows: an event header, Monte Carlo specific ('truth') information like generated 4-vectors and hits, and simulated digitized raw data. Such a Monte Carlo file is preceded by data flows describing the Monte Carlo version and parameters. A real measured event consists typically of a header, raw digitized data, trigger data, and information on the status of the DAQ.

5.4.2 Detector Geometry Description

A convenient way of describing detector geometry is implemented in GEANT [54]. In order to make a convenient detector description that can be handled by all ZEUS offline software, this structure has been ADAMOized [63].

In GEANT, one iteratively positions specific volumes within mother volumes, which can be placed into grandmother volumes, and so on. A position is characterized by a position vector in the mother coordinate frame, and a rotation matrix. A volume has a tracking medium, consisting of a specific material, or a mixture of materials. The material of a daughter volume overrules the material of the mother volume in which it is located. In this way, one can easily set up complicated structures and make sure that unoccupied spaces are filled with the desired material, such as air.

The GEANT structure is a typical example of a heavily interrelated set of entities, as proved by the frequent use of the ZEBRA link system, and as such its logical structure is conveniently described in ADAMO. The resulting detector geometry description structure is used by all components.

5.4.3 Monte Carlo event generator output

A standard, consistent data structure has been set up in order to handle the output of various Monte Carlo event generators on the 4-vector level [64]. This structure allows a transparent view of the generated event, and data in this structure can be fed as input to the detector simulation program. The structure is described in figure 5.2, and the information in it is fully compatible with the standard HEPEVT common block [65]. The ZDIS program provides its output in this format.

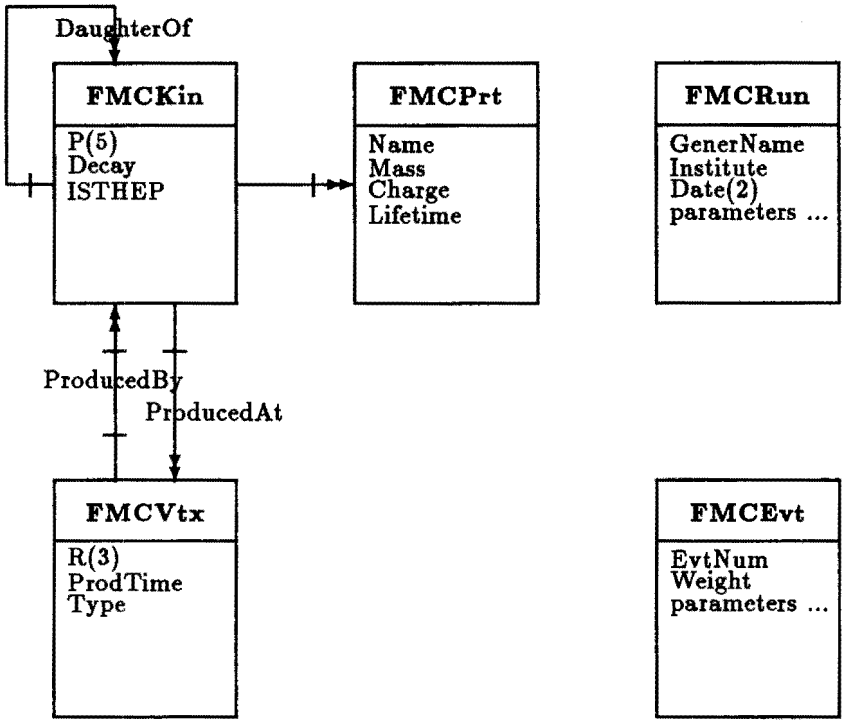

Figure 5.2: *The ADAMO structure of the event generator output.*

5.4.4 Calibration and conditions constants

The components calibration data is also stored in GAF format in the calibration constants base. The validity range of a given set of calibration constants is stored in the key of the corresponding data flow. Some components need additional information on the running conditions, which are handled similarly.

5.5 The Detector Simulation Program MOZART

After the Monte Carlo generation of the event 4-vectors, the second phase of Monte Carlo simulations usually consists of the simulation of the response of the detector to these events, as if they were real events. For the ZEUS detector, this is performed by the detector simulation program MOZART (MOnte carlo for Zeus Analysis Recon-

struction and Triggering). MOZART is based on the general detector simulation package GEANT. MOZART accepts as input the output of ZDIS, whose logical structure is shown in figure 5.2. Accepted 4-vectors are tracked through the detector, adding hits in sensitive detectors, and digitizing them. As a result, MOZART adds a data flow to the data, containing the hits and digitizations, which is input to the reconstruction program.

It has been shown that the simulation of hadronic showers in GEANT does not describe the ZEUS test data measurements [66]. For this reason, as well as for improved speed, the showering routines were adapted and modified in a major way [66]. *Shower terminators* were introduced, that cut short the GEANT tracking and showering procedures when the energy of a particle falls below a certain cut-off, and distribute its energy according to a parametrization following from the test results. In this way, MOZART agrees with the test results, at the cost of losing its 'first principles' predictive power. In addition, the detailed uranium/scintillator layer structure was replaced by a uranium/scintillator *porridge* with effective identical properties.

MOZART has the option of simulating uranium and electronics noise of the calorimeter, by adding to the calorimeter data energy generated according to a convenient parametrization.

5.5.1 The Data Banks for the Uranium Calorimeter

The aim of MOZART is to output the data in the same format as expected from the detector. The raw data tables for the calorimeter are named CFTENE, CBTENE and CRTENE, and are described in section 6.2. The structure of the output of MOZART for the calorimeter is shown in figure 5.3. This output consists of 'truth' information. i.e. the hits and their relation to the generated 4-vectors, as well as raw digitized information in the CXTENE tables.

The generated 4-vectors are stored in table FMCKin, as shown in figure 5.2, and are treated as tracks by MOZART. The FMCFte table contains information on the fate of a given track, and its entry point and direction in the calorimeter. The CcGHit table stores per track and per PMT the energy deposited by GEANT in scintillator and wavelength shifter, and is in addition linked to the resulting raw digitized data tables CXTENE.

5.6 Trigger Simulation

The ZGANA program takes care of the offline simulation of the ZEUS trigger logic. This can then be used in the design of trigger criteria, in the evaluation of the actual trigger performance, and in calculating the trigger acceptance. The ZGANA software library is set up as a slave system, and consists of a large number of routines to simulate parts of the trigger system. This modular set up makes it possible to simulate in a flexible way various combinations of trigger logic. ZGANA

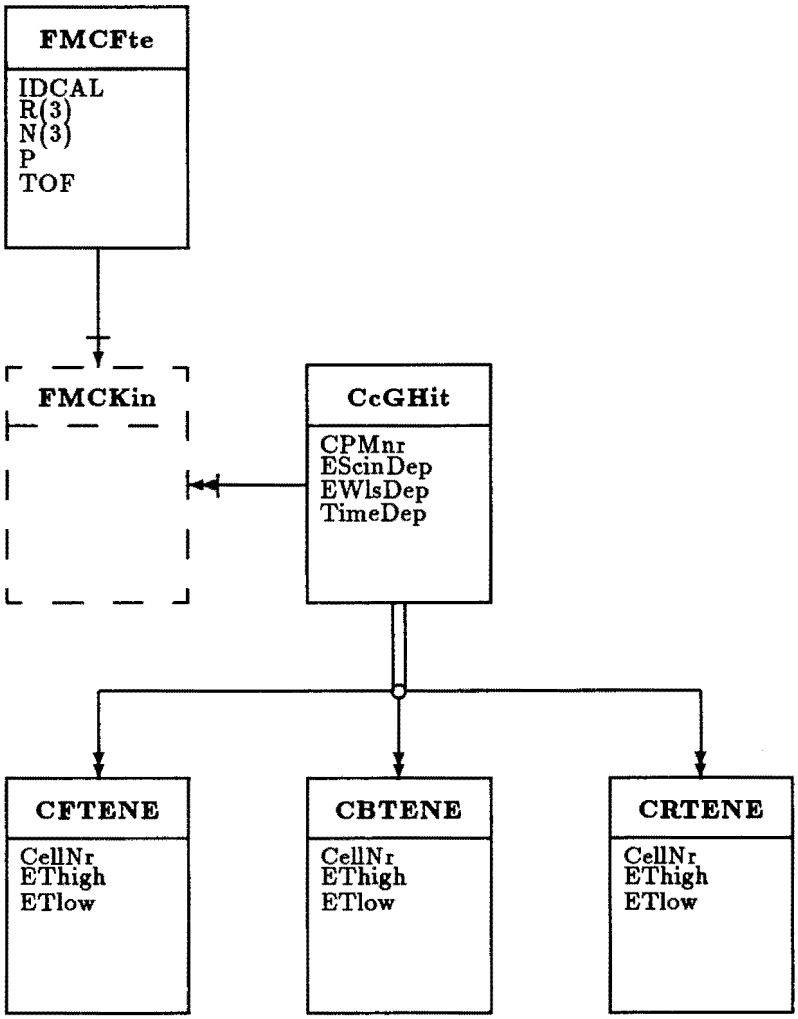

Figure 5.3: *The MOZART output data structure for simulated calorimeter data.*

reads output of MOZART as well as real events, and adds a data flow containing the information of the trigger simulation.

5.7 The Reconstruction Program ZEPHYR

The ZEPHYR (ZEus PHYsics Reconstruction) program uses the raw measured event data, or digitizations, to reconstruct objects that are to be used in a subsequent physics analysis. For the tracking detectors, such objects are track elements and tracks, and for the calorimeters and HES calibrated energies and energy clusters.

ZEPHYR consists of three major phases. In phase 1 of the program, each detector component module operates on its own component-specific data, in order to obtain phase 1 reconstructed objects, such as local track segments or clusters. In phase 2, these objects are matched and linked to each other over detector boundaries. The results of this procedure can then be used again in a reevaluation of the component-specific objects. Finally, a global reconstruction is performed, which can result in fully reconstructed and identified tracks, and one or more reconstructed vertices.

The reconstruction procedure for the uranium calorimeter will be described in detail in the next chapter. In this section we describe some of the other relevant parts of the reconstruction.

ZEPHYR reads as input the data flows containing the digitized detector data, either from a real event, or from a simulated event. In addition, for real data only, the information on the status of the DAQ, is used. The data from the trigger processors is normally not used, although in principle it could. Normally, the data is stored on cartridges on the main DESY IBM. The actual reconstruction is performed on a farm of powerful Silicon Graphics multiprocessor workstations. A server controls the raw and reconstructed data flowing between these machines, and the distribution of events to the individual processors. A first reconstruction pass is performed usually within a few days after data taking, and is accurate enough to allow a first selection of a physics data sample. When more accurate calibration data is available, a second pass is performed, whose results are used in the physics analysis. The output of the reconstruction program is added to the event data in the form of a new dataflow.

5.7.1 Context diagram

A context diagram defines a system in terms of its environment, and the nature of its relations with the outside world. In figure 5.4 the context diagram for the reconstruction program is drawn. ZEPHYR accepts simulated as well as real data. The time at which the event is taken is part of the event data, and guides the selection of geometry and calibration constants.

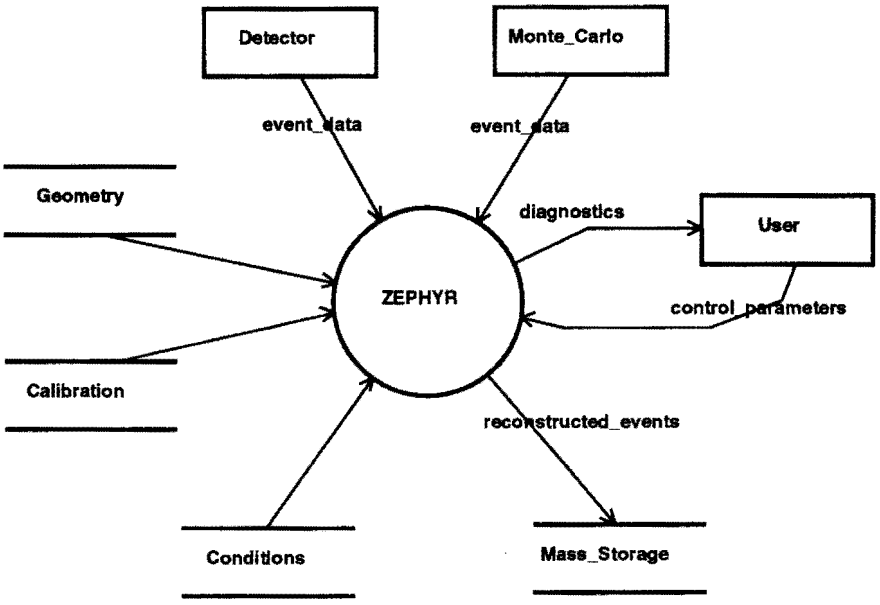

Figure 5.4: *ZEPHYR Context Diagram.*

5.7.2 Components of the Program

In this section the various modules in the program are shortly described. Since some detector components were not, or only partly, equipped with readout or trigger electronics in the 1992 data taking period, the quality of reconstructed objects for these components did not meet the design criteria in 1992. The reconstruction efficiency and accuracy for these components will therefore be described in section 8.2. The various phases of ZEPHYR are schematically shown in figure 5.5. Phase 1 is performed in parallel for most components, except for the HES, that needs information from CTD and CAL.

Track Reconstruction

Phase 1 track reconstruction is normally performed in VXD, CTD, FTD and RTD, and in the muon chambers. Since FTD and RTD did not take part in the data taking in 1992, the corresponding modules in ZEPHYR were turned off. The track reconstruction code is able to find tracks using standard techniques of seed finding, pattern recognition and track fitting. A minimum of 4 hits is required to create a track. Optionally, hits recorded in the VXD can be used in the CTD track reconstruction software as if they were recorded in an additional superlayer of the CTD, note that the VXD has a better position resolution than the CTD.

The efficiency of track finding within the region of good acceptance has been

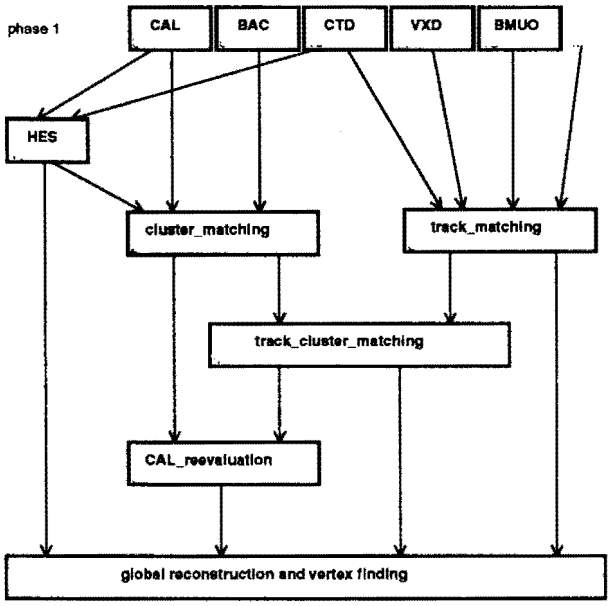

Figure 5.5: *The various phases of ZEPHYR.*

shown to be high. It should be noted, however, that many tracks at HERA have low p_T and do not, or only partially, enter the innermost superlayer of the CTD.

An alternative track finder, as used in the third level trigger, reaches compatible accuracy.

The global track matching procedures in phase 2 of ZEPHYR, extrapolate VXD, CTD or muon chamber tracks over detector boundaries, matches them with each other, and refits them for better accuracy.

The Uranium and Backing Calorimeters

The reconstruction code for the CAL is described in the next chapter. The backing calorimeter measures particles penetrating through, or produced in, the main calorimeter. The clusters found in the phase 1 reconstruction procedure in the BAC have a very limited energy and position resolution. They are classified as either 'jet', which are large clusters presumably caused by energy leaking out the back of the main calorimeter, originating from the showers produced by high energy jets that are not fully contained in the CAL (see section 4.4), or 'mip', which are muon candidates that penetrate the calorimeter as a minimum ionizing particle. In phase 2, the phase 1 objects from CAL and BAC are matched in position, without using information from any other detector, and the energies are added. In the energy addition, corrections are made for energy loss in the dead material, such as calorimeter

support structures and cables, between the active parts of the two detectors [67].

HES

The reconstruction of clusters in the hadron-electron separator HES uses tracks and clusters that were reconstructed in the CTD and CAL in phase 1 of ZEPHYR. The tracks are extrapolated into the HES, and matched with energy deposits. If additional energy deposits are found that do not match with tracks, 'neutral' clusters are formed. The location of the HES in the calorimeter ensures large differences between the signals of hadrons and those of electrons or photons. In the 1992 data taking, the HES objects have only been used in the final analysis as an independent check on reconstructed electrons in the CAL. In the future, it is likely that they will be used in the phase 2 cluster matching, and in the phase 2 reevaluation of calorimeter data.

Track-cluster matching

Reconstructed tracks are extrapolated through the main calorimeter, either using the accurate, but slow, GEANT tracking package GEANE, or using a faster parametrization. From the list of traversed cells, a link to reconstructed clusters in the calorimeters can be made.

Global Reconstruction and Vertex Finding

The term global reconstruction, or phase 3, refers to the reconstruction of vertices and, where possible, 4-vectors created in those vertices, by combining the information of all components. Due to the limited information provided by the tracking detectors, this included only a global primary vertex fit in the 1992 data taking period.

The accuracy of the vertex reconstruction depends heavily on the number of tracks participating in the fit, and the quality of the tracks. A good vertex reconstruction requires at least two tracks, although in principle in events with a single track, the position of closest approach to the beam axis could be used.

The most interesting parameter of the vertex is in first instance its position in z, since this translates directly into the angle of scattering of the electron in a NC event, and can be used to validate the hypothesis that we have ep collisions, and not beam-gas events. The reconstructed position in the $x - y$ plane can be used to identify events with a vertex in the beam pipe wall, but these are less frequent. The vertex reconstruction accuracy in the z coordinate is considerably poorer than in the $x - y$ plane.

Also the TLT tracking code provides a vertex fit, with similar results.

5.8 Analysis packages

A framework (EAZE) has been set up in order to facilitate the handling of ZEUS events in a physics analysis application. Since the GAF format is used for all ZEUS data, all data flowing between the various offline programs can be handled by EAZE. A library of physics analysis routines, is available for common tasks like electron finding, kinematics reconstruction, jet finding and Lorentz boosts. In addition, interactive packages are available for data scanning and analysis.

Chapter 6

Calorimeter Reconstruction Software

In this chapter, we will describe the design and implementation of the reconstruction software for the uranium calorimeter [68].

We will first show the structure of the program, describe the format of the raw data, and the description of the geometry of the calorimeter. We will then show how the best estimate of the energy deposited in the cells is obtained, using the calibration data. We will discuss noise levels and noise suppression cuts. The local and global clustering algorithms are explained, as well as the methods for particle identification in the calorimeter. We will show the accuracy and efficiency of the reconstruction of electromagnetic showers (electrons and photons) in detail, and we will evaluate the implemented jet algorithms. Finally we will shortly discuss the difference between clusters reconstructed off-line and clusters found by the trigger processors.

Where convenient, we will use data flow diagrams (DFD's) and entity-relationship diagrams (ERD's) to illustrate the program and the data.

6.1 Structure of the Program

The calorimeter reconstruction software is incorporated within the framework of the ZEUS reconstruction program ZEPHYR. Therefore, the context diagram for ZEPHYR, drawn in figure 5.4, also holds for the calorimeter reconstruction program. In particular, the program must be able to handle real and simulated data on the same footing, be able to read from the geometry and the calibration constants base, and keep track of user-supplied parameters overruling the default values.

The reconstruction is performed starting from the data flow containing raw digitized event data. Using the information on the geometrical structure of the calorimeter, and calibration data, clusters of energy are reconstructed. These are

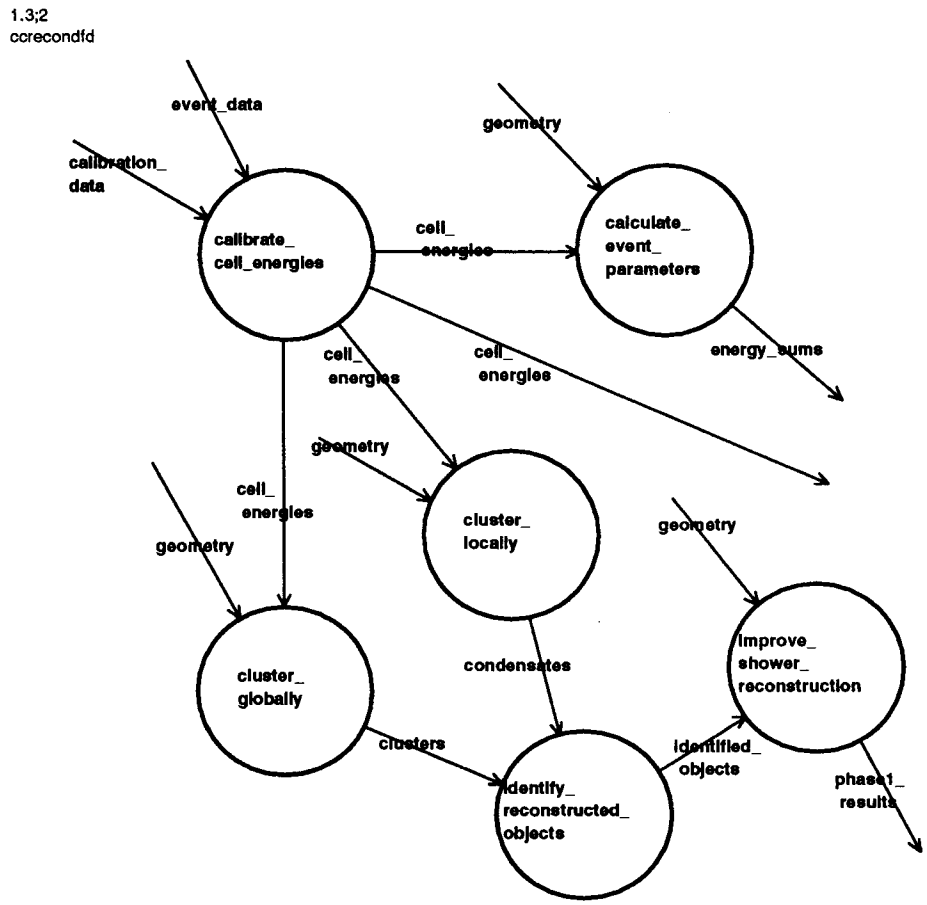

Figure 6.1: *Top level data flow diagram of the calorimeter reconstruction software.*

subsequently matched with information of other components, as described in section 5.7 and figure 5.5. The output of the reconstruction program is added to the event data in the form of a new data flow.

We can expand the context diagram into a top level data flow diagram for phase 1 of the calorimeter reconstruction, as shown in figure 6.1. Such a top level DFD describes the general structure of the program in terms of information flow and data processors.

In phase 1 of the reconstruction, the calorimeter software makes use of the calorimeter data only, and performs the following tasks:

- the (re)calibration of the energies in the cells
- the local clustering into *condensates*

- the global clustering in the calorimeter into jet-like *clusters*
- the identification of clusters and condensates
- improved reconstruction of identified clusters and condensates
- the calculation of calorimeter-specific event parameters, such as energy sums.

As shown in figure 5.5, clusters and condensates are subsequently matched in phase 2 with reconstructed clusters in the backing calorimeter, and the resulting objects are matched with tracks. The HES reconstruction program uses phase 1 reconstructed calorimeter clusters and CTD tracks in an evaluation of HES data. The reconstructed HES clusters are used in the final phase of global reconstruction, but will in the future probably be added to the CAL and BAC clusters in the cluster merge procedure in phase 2 of ZEPHYR. The calorimeter has the option of a reevaluation of calorimeter data after the global cluster and track-cluster matching has been performed; at the moment this consists of only a small modification to the condensates.

6.2 Raw calorimeter data

When the trigger system accepts an event, all 11836 channels connected to photomultiplier tubes are read out, and their data is passed to the off-line mass storage, as input to the reconstruction. The basic information of a channel consists of an energy and a time.

The readout electronics, and the measurement of the deposited energy and the time of the pulse in a channel, are described in section 4.4.2. Under normal data taking conditions, the DSP's will reconstruct the energy and the time in a channel following equations 4.15 to 4.18, and pass the values of these quantities in MeV or ns. However, it might happen that a DSP decides, for whatever reasons, that it cannot accurately reconstruct energy and time for a specific channel. In these special cases, the samples h_i will be written out for subsequent off-line analysis. The reconstruction program is able to handle both formats, the choice is driven by the data itself.

The precision of the measurement of reconstructed energy and time is such that they can be packed into one 32 bit word, with the time occupying the 8 most significant bits. As described in section 4.4, a calorimeter cell is always read out by a pair of 2 photomultiplier tubes, which are connected to the wavelength shifter bars of either the left or the right side of a calorimeter module. These two PMT's of a cell are labeled *high* and *low*, referring to their values of x (FCAL,RCAL) or ϕ (BCAL) in the global ZEUS coordinate system. In the data, such a cell will occupy one row in the raw data ADAMO tables CFTENE, CBTENE or CRTENE (see figure 5.3), whose columns are:

1. **CellNr**: a unique number characterizing the cell;

2. **EThigh**: data from the high x or ϕ PMT;

3. **ETlow**: data from the low x or ϕ PMT.

The connection to the cell position is made via the cell number.

Monte Carlo data passed through the detector simulation is delivered in the same format, as shown in figure 5.3.

Since the 1992 data taking period was the first data taking period under actual HERA conditions, all channels were read out in every event. In this way, the energy flow in real *ep* interactions as well as in background events could be studied, and zero-suppression schemes investigated. The price to be paid in the described format is a lengthy 80 kByte raw calorimeter data per event. Two alternatives are being investigated for future data taking: a further compression of the data of all channels by using a more sophisticated data packing scheme, and *zero-suppression*, i.e. only passing channels with energy above some predefined cut to the output. The latter scheme will imply a large reduction of the amount of channels to be treated by the reconstruction program, but may influence the calorimetric measurement of event energies.

Other tables passed to the reconstruction in each event contain various information on the status of the readout electronics and the individual channels; this information is used in the treatment of bad channels as described in section 6.7. Additionally, a data flow exists that contains the output of the trigger processors. Of special interest are the clusters and energy sums calculated in the CAL-SLT.

6.3 Calorimeter Geometry Description

The full geometry of the calorimeter is stored for simulation and reconstruction purposes in the detector geometry description scheme described in section 5.4.2. As has been mentioned, this scheme is essentially an implementation in ADAMO of the full GEANT way of setting up a detector geometry.

For the calorimeter, three (dummy) volumes, FCAL, BCAL and RCAL, are positioned in the mother volume HERA. The build up of the elements within RCAL is very similar to the one in FCAL, but when positioning RCAL in the mother volume HERA, it is rotated by 180°. Within FCAL, the inactive material of endbeams, C-arms, support structures, lead sheets, and photomultiplier(shielding)s is modeled. The active volume of FCAL is filled up by repeated positioning of cell-like BOX volumes. In order to increase the speed of the simulation of showers in GEANT, the uranium/scintillator sandwich structure is modeled as a single *porridge* of uranium, steel, paper and plastic. Each cell-like BOX, which can be subdivided (GEANT *division*) into 2 or 4 EMC cells, contains appropriate daughter volumes for the wavelength shifters, the silicon gap(s), and the porridge.

BCAL is filled up by 32 daughter volumes which, though called TOWE, actually represent the modules, appropriately positioned and rotated. In turn, each TOWE again is filled by the appropriate number of BEMC, BHAC1 and BHAC2 cells.

The cells are no longer simple boxes, but polygons (PGONs), in order to model the (partially) projective structure of BCAL. Each cell has a structure similar to the FCAL cells with wavelength shifter, silicon gap, and porridge.

Additional volumes, unrelated to the calorimeter itself, model various detector support structures.

The calorimeter reconstruction program in ZEPHYR does not use the described GEANT structure directly. It is in first instance interested in the actual physical cells and their locations in the global ZEUS coordinate system. Each cell is uniquely labeled by its cell number, which is easily obtained from, and decoded into, module and tower number, and which is also contained in the raw event data. At initialization time, the reconstruction program derives from the full GEANT structure a new table CUCELL, containing the position of the centre of gravity of all 5918 calorimeter cells. An Entity Relationship Diagram is shown in figure 6.2. The CUCELL table is linked with tables from the original GEANT geometry structure (dashed contours), so that when other parameters of the cells are needed, such as size or rotation, they can easily be looked up in the appropriate tables.

Additional useful information of a cell is given by its *neighbourhood*, i.e. its physical configuration with respect to other surrounding cells. These cell configurations are stored in an *adjacency* table CCAdJa. For each cell, we list the neighbouring cells with which it shares at least a part of its side. We process this list by subtracting, from the actual cell numbers of the neighbours, the cell number of the original cell; the resulting list of cell number differences defines an *adjacency class.* Cells with similar configurations (cells in central parts of the calorimeter, or bulk cells), fall in the same adjacency class. Cells at the edges of the calorimeter (surface cells) generally differ. An intermediate table CUAdCl holds the relation between a cell and its adjacency class. The cell adjacencies are used in the local clustering procedure. No attempt has been made to connect cells over subdetector (FCAL-BCAL, or BCAL-RCAL) boundaries.

6.4 Cell Energy and Time Calibration

The calibration of the energy measurement is of crucial importance for a calorimeter. This includes setting the energy scale and measuring the linearity and uniformity of the calorimeter response before operation, as well as constantly monitoring the stability of the response during operation. The methods to do this for the ZEUS calorimeter have been described in section 4.4.3.

The energy and time are calculated for all channels on the DSP's, as described in section 4.4.2. In order to do this, the DSP's have access to the vast amount of constants that characterize the electronics (see section 4.4.3). Neither these constants, nor the original raw data on which the constants work, are available in the off-line reconstruction, and the calculation can therefore not be redone. For this

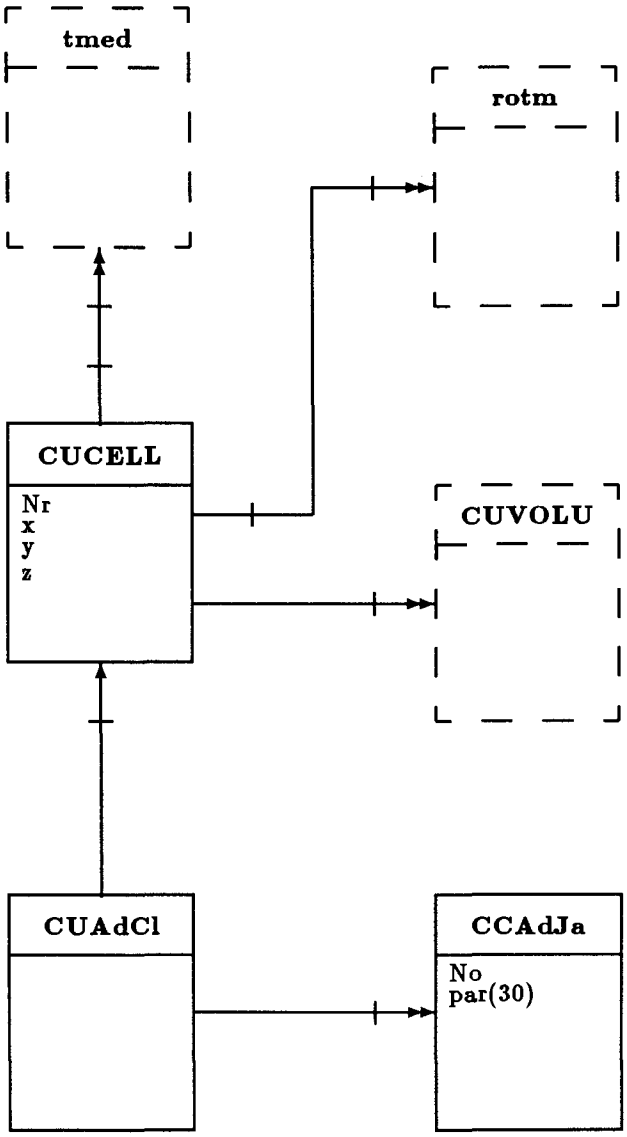

Figure 6.2: *The* CUCELL *geometry data structure*

reason, the constants as well as the actual reconstructed channel energies per event are continuously carefully monitored on-line.

The DSP's send their results in units of MeV c.q. ns. Several effects can influence the accuracy of the measurement, and the DSP's cannot correct for all of them. Effects to be repaired offline include:

- variations in the PMT gain. Since this effect cancels in the ratio $\frac{particlesignal}{UNO}$, the uranium noise can be used to correct for it. The PMT gain, which determines the energy scale of the calorimeter, is set initially in such a way that the uranium current equals a nominal value I_{UNO}^{nom}. During data taking, this current is remeasured regularly. Let the latest measurement of the uranium current in a channel be I_{UNO}^{meas}, and the energy reconstructed by the DSP's in that channel in a given event be E_{DSP}. The corrected energy in that channel in that event can then be expressed as:

$$E_{cor} = \frac{I_{UNO}^{nom}}{I_{UNO}^{meas}} E_{DSP} \tag{6.1}$$

- non-linearities in the PMT response at high energies.
- detector non-uniformities and inefficiencies.
- variations in the offsets to the measured time. Each channel has an individual offset that needs to be taken into account in the time reconstruction. These offsets are related to effects such as different wavelength shifter and light guides lengths, and different transit times in PMT's and electronics. In addition, the geometrical location of the cell (i.e. the time of flight of a particle between vertex and calorimeter) plays a role. The offsets are chosen such that all channels with energy from an actual *ep* interaction in the nominal interaction point give $t = 0$. The offsets can be calculated with laser calibration runs. In addition, the timing of the whole calorimeter is influenced by different timings of the beams, which may vary from run to run.

Calibration constants are obtained from the analysis of special calibration events, which may be taken in separate calibration runs, or be obtained by test triggers during normal data taking. Such test events include the measurement of the uranium noise and the injection of a well known amount of charge in the front end electronics. Eventually, the calculated calibration constants are stored in ADAMO format in GAFs in the constants base, where they can be read by the reconstruction program.

In the reconstruction of the 1992 data sample, only the energy correction with the UNO, and the offset correction to the time were implemented. These are in fact the dominant corrections, the other corrections are expected to be much smaller. In the 1992 data taking period, UNO calibration data was typically renewed every 3 days. This is sufficient to compensate for the long term fluctuations in the PMT gain, and ensures the energy scale to be correct to within 1%, as required.

Output Table

The recalibrated cell energies are available in ADAMO table Caltru. Instead of storing the energies of the 2 PMT's of a cell separately, we store the total energy in the cell, and the *imbalance*, representing the difference between the two channels. The output tables are described in more detail in section 6.16. An entity-relationship diagram of the structure of the output tables is shown in figure 6.21.

The reconstruction program calculates some general event parameters like total energy and missing energy, by looping over all cells in Caltru. The result is stored in table CR1obj.

6.5 Noise and Pedestals

Two sources contribute to a non-zero energy in the calorimeter cells in empty events: noise originating from the electronics, and uranium noise (UNO). The uranium noise in normal events in data taking at HERA differs from the uranium noise signal used for calibration purposes as described in section 4.4.3. For calibration purposes, the UNO signal is integrated over a relatively large time of 20 ms, but in data taking at HERA, the effective gate time is $\mathcal{O}(100)$ ns, and the UNO forms a background to the signal. In normal events, the mean noise signal is not recorded, but subtracted from the measured cell energy as a pedestal. The noise contributing to the measured energy is therefore caused by the fluctuations in the uranium signal.

In channels with little energy (less than $\mathcal{O}(10)$ GeV), the high gain mode is used in the determination of energy and time. In normal *ep* events, this is the case for almost all cells. Therefore, the noise in this mode has the largest influence on the measured energy. Noise in the low gain mode contributes to an already much larger signal. The UNO is in fact the dominating source of noise in the calorimeter: when the high voltage on the PMT's is lowered, the uranium noise signal is no longer amplified and the total noise is dominated by the electronics noise; this total noise turns out to be much smaller than the noise at full high voltage.

If the constants used by the DSP's in the energy calculation are correct, the mean energy of a cell, and thus of the whole calorimeter, should be zero for empty events. The width of the energy distribution in the cells in such events is determined by the noise, and the noise should give equal positive and negative contributions to the energy.

Figure 6.3 shows the total energy in FCAL, BCAL, RCAL and in the whole calorimeter, per event, in real events. While this data was taken, a proton beam circulated in HERA. This ensures that the circumstances under which this data was taken are realistic for normal data taking conditions at HERA. The contributions to these plots from beam-gas events or cosmic muons is negligible, since almost all of these events were triggered by sparks (see section 6.7), and the signals from the sparking PMT's were removed from the cell list.

Figure 6.3 shows that the total energy in these events, summed over all 11836

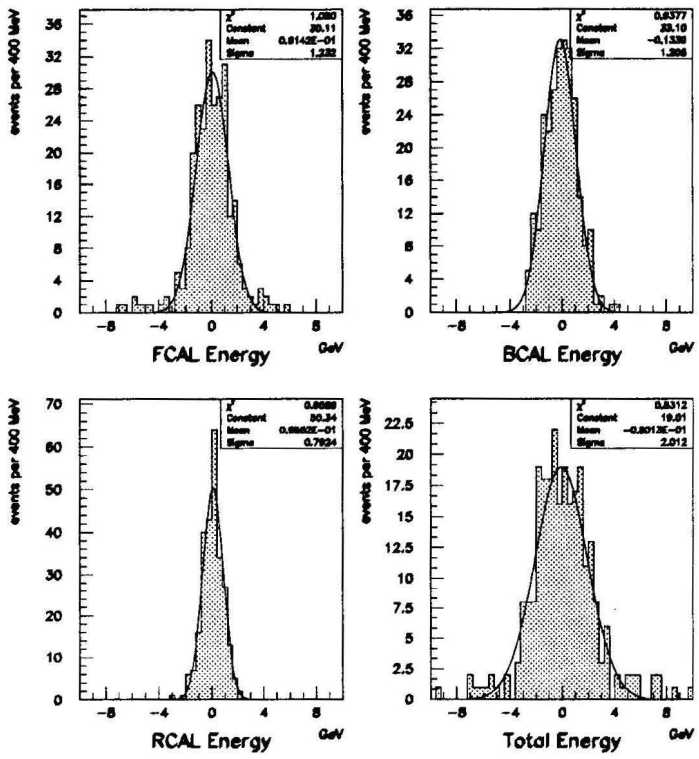

Figure 6.3: *The total energy measured in FCAL, BCAL, RCAL, and in the whole calorimeter, per event, in data taken with spark triggers. The sparking PMT was removed from the cell list, these events are therefore empty.*

channels (excluding the sparking PMT), differs on average less than 100 MeV from zero, whereas this distribution has a width of 2 GeV. We conclude that the constants have been determined with high enough accuracy to ensure an 'empty' calorimeter in empty events.

Figure 6.4 shows, for the same events, the distributions of the energy measured in *individual* FEMC, FHAC, BEMC, BHAC, REMC and RHAC cells. The distributions are centred around zero, as expected, and the width is determined by the noise levels. The data shown here is typical for the noise data that is continuously taken in order to monitor the calorimeter. The average measured noise levels, i.e. the average widths of the distributions of the energies in the cells, are 18 MeV for FEMC and REMC cells, 15 MeV for BEMC cells, 26 MeV for FHAC cells, 25 MeV for RHAC cells, 28 MeV for BHAC1 cells, and 32 MeV for BHAC2 cells. The contribution from the electronics noise to these numbers is typically a few MeV.

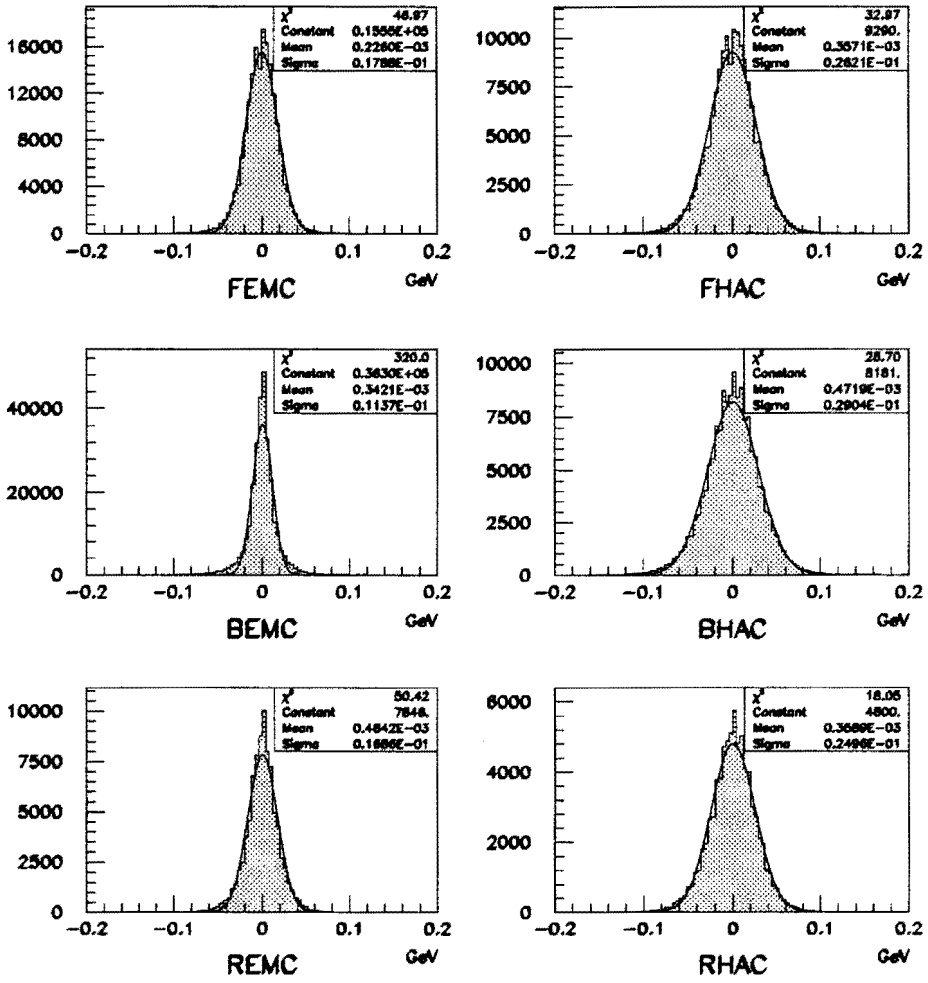

Figure 6.4: *The energy measured in various kinds of calorimeter cells, per event, in data taken with spark triggers. The sparking PMT was removed from the cell list, these energies are therefore due to noise.*

6.6 Cuts in Cell Energies

In the filling of the table Caltru, which contains the energy measured in the individual cells, it is possible to include only cells with energy above a predefined cut. There are various reasons why this is convenient:

- Cells which have no signal but only noise are suppressed;

- The cell table is reduced in size, which is convenient for the subsequent clustering procedures;
- If the constants used in the energy calculation are not quite correct, each cell has a mean energy that slightly deviates from zero (a *pedestal*). When summing energies over all calorimeter cells, this can introduce a considerable bias.

However, by applying a cut on the energies in the cells, part of the signal from actual particle showers may also be lost. In particular, hadronic showers, with their large transverse size, can spread over many cells. In the tail of these showers, the cells have little energy, but are so numerous that they can contribute significantly to the total energy. In addition, a cut on the energies of the cells implies that the high positive tail of the noise distribution survives the cuts, but is no longer balanced by a similar negative tail of the noise distribution, since that has been cut away.

The cuts in cell energies discussed here are cuts made in the off-line reconstruction. It is in principle also possible to make such cuts on-line in the data acquisition: in that case only cells with energy above the cut are read out and stored.

The consequences of cuts in cell energies on various event related quantities have been investigated on real data. This data consisted of a subset of selected DIS events taken in 1992, whose selection will be described in chapter 8. The results are plotted in figure 6.5. The average total energy in the calorimeter in these events equals 68.28 GeV, when summing over all 5918 calorimeter cells, and without application of any energy cut. Figure 6.5 shows the average number of cells per event surviving the cell energy cut (upper left), the average total energy in the calorimeter per event (upper right), the average time needed per event to UNO-calibrate the data and fill Caltru (lower left), and the average time per event needed to perform the local clustering (lower right), as will be explained in section 6.8. The time is plotted in units of CPU seconds on a DEC-5000 computer, but the absolute numbers are irrelevant in this context. On the horizontal axes of the figures, we plot the value of the cell energy cut that has been applied. The shown values represent the value of the energy cut on the EMC cells; in all cases the corresponding cut on the HAC cells equals twice the value of the cut on the EMC cells. The mentioned cuts apply to all EMC and HAC cells, in FCAL, BCAL and RCAL. Values of the plotted quantities obtained without energy cuts, are written on the plots, and in addition shown as a solid line in the total energy plot. An example to illustrate the plots: when no cuts are applied, all 5918 cells are accepted. When a cut of 0 MeV is applied on EMC and HAC cells, approximately 3000 cells are left (as expected). When a cut of 50 MeV is applied on the EMC cells and 100 MeV on the HAC cells, only slightly more than 100 cells survive the cuts.

One may conclude that the number of cells surviving the energy cuts falls off very steeply, as expected from the noise spectrum. When all negative cell energies are cut away by a cut of 0 MeV, the average total energy in the calorimeter is about 45 GeV higher than when no cut is applied, since the positive part of the noise is no longer balanced by the negative part. When the cut on the cell energies is

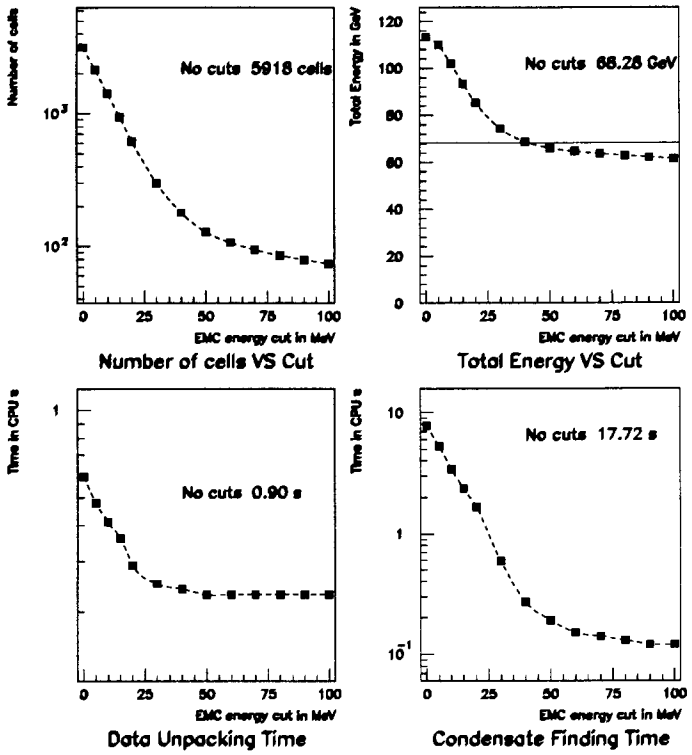

Figure 6.5: *Influence of cuts in cell energies on various quantities in real events. For a description of the plots, see the text.*

increased, the average total energy drops, reaching a value of 68 GeV at an energy cut on the EMC cells of 40 MeV (i.e. 80 MeV on the HAC cells). The time needed for data unpacking and calibration, and in particular for clustering, is very sensitive to the number of cells surviving the cut, and therefore to the value of the cut itself.

In view of figure 6.5 and the average noise levels that have been described in the previous section, it has been decided to apply a standard energy cut of 60 MeV on EMC cells, and of 110 MeV on HAC cells (100 MeV on HAC0 cells) in the reconstruction. Eventually, the energy of reconstructed clusters must be corrected for the energy lost in cells that do not survive the cuts.

6.7 Bad channels and sparks

The uranium calorimeter is constantly carefully monitored for bad channels, where the reconstructed energy cannot be trusted. It is required to keep the number of bad channels below 1 or 2 % of the total number in order not to influence the

calorimetric energy measurement. Channels can be recognized as bad during one of the standard calibration runs, or via an abnormal UNO value. Such channels are included in a list, which is sent as a bit pattern with the data of every event. The DSP's also use this list, and set the energy of a bad channel to 0.

Each cell in the calorimeter is read out by two PMT's. The two PMT's of a cell are connected to different front end cards. When a front end card fails, only the channels on one single side of a module are affected. In this way, a certain amount of redundancy is built into the system, since it seldom occurs that both channels of a cell are bad (in which case the cell is known as a *hole*).

In the reconstruction, these bad channels are recognized via the bad channel bit pattern that is sent with the event data. In the Caltru table, a correction is made for cells with a bad channel. The total energy in the cell is recalculated by doubling the energy of the good channel, and neglecting the bad channel. The imbalance, or the difference between the two channels, is set to 0. This procedure is equivalent to equaling the energy of the bad channel to the energy of the good channel. It will result in a loss of accuracy in the reconstructed energy, since only the energy sum of both channels is, in first order, insensitive to the point of impact of the particle that created the shower. It will also result in a loss in accuracy of the position reconstruction of electromagnetic showers, since the difference between the two channels is used, as explained in section 6.12.

It has been observed (see section 8.2), that often a single channel in the calorimeter shows a considerable amount of energy, unrelated with beam activity, and at a random time. These are thought to be caused by discharges in the PMT base or between PMT and shielding, due to the build up of a static charge at the photocathode of the PMT, in particular in BEMC cells. Such *sparks* can be recognized by the large imbalance of the cell energy, and the low energy in the remaining cells of the calorimeter, and subsequently be removed from the analysis.

6.8 Local Clustering and Condensates

The first reconstruction step after filling the cell table Caltru, is *local* clustering. It is the aim of the local clustering procedure to gather the cells that are hit by a shower produced by one, or as few as possible, single particle(s). The resulting reconstructed object should give the energy deposited by that particle, as well as the position of the centre of gravity of its shower in the calorimeter. We have chosen to adopt the name *condensate* for such an object, as the name *cluster* will be used for the output of a jet algorithm running in the *global* clustering, described in the next section.

The local clustering starts off from the cell energies as stored in table Caltru. The clustering is performed on a simple geometrical basis, using the cell neighbourhoods, or adjacencies, as described in section 6.3. The procedure is as follows:

1. Cells with energy above some user-defined cut are placed in a list, ordered on

energy.

2. The first cell, with the highest energy, serves as the seed for a condensate. Using the adjacency table, all neighbours of the seed cell are scanned for energy. If a neighbour has enough energy , it is added to the condensate. The neighbour cell will be inserted in two selectors [1], one denoting that this particular cell has been used in a condensate, and the other denoting that, in turn, the neighbours of this cell need to be scanned.

3. Using the second selector, and the adjacency table again, the neighbours of the original neighbour cells are scanned for energy, taking care that cells already used (selector 1) are not used again. If these cells have enough energy, they are added to the condensate and inserted in both selectors. When all neighbours of a particular cell have been scanned, it is deleted from the second selector.

4. This procedure is repeated until the second selector, of cells whose neighbours remain to be scanned, is empty. The condensate is now complete.

5. If there are still cells in the list that have not yet been used in a condensate, the cell with the highest energy of these serves as the seed for a new condensate, and the whole procedure is repeated.

6. The steps above are repeated until all cells in the list have been used. An additional energy cut is then applied to the found condensates, and condensates that do not pass the cut are deleted from the table.

The resulting condensates are objects of energy deposits spread over a number of neighbouring cells. The energy of a condensate is the sum of the energy of the cells assigned to it. Its reconstructed position is the energy weighted average of the centres of those cells. Where appropriate, the direction of the condensate in the calorimeter is determined from the condensate centres in EMC and HAC.

The definition of the neighbourhood of a cell is chosen such that for all cells, the adjacent cells that are scanned for energy are those cells with which the original cell shares at least a part of a side. That includes the cells above, below, on the left of, on the right of, in front of, and behind the original cell, but not the cells diagonally above or below it. The argument for not using the diagonal cells is the fact that this would considerably increase the number of neighbours to be scanned for each cell, and since the scanning procedure is an iterative process (of every accepted neighbour, the neighbours need to be scanned), the CPU time needed for the clustering would increase dramatically. For most showers, these diagonal neighbours are accessed in any way as the left or right neighbour of an upper or lower neighbour of the original cell. For showers where this is not the case, for example because there is not enough energy in the non-diagonal neighbours, the resulting incompleteness of the condensate can always be resolved in phase 2, by

[1] A *selector* contains a subset of selected rows from an ADAMO table, and provides a fast and flexible access method to those rows.

reanalyzing the condensates, and merging incomplete ones with other condensates. This also holds for the consequences of the fact that the neighbourhood definition does not cross the FCAL-BCAL and BCAL-RCAL boundaries, as the concept of neighbourhood in these regions is very hard to define.

It must be noted that the local clustering into condensates is completely independent of the particle type that actually caused the energy deposit. This makes condensates useful for all particle types, and does not introduce assumptions in the early stage of the reconstruction. However, it is clear that many corrections and improvements are possible that do depend on the particle type. The shower position reconstruction for electrons, for example, can be improved using the individual energies recorded by the two PMT's of a cell. The reconstructed energy can be corrected for several position dependent effects, such as those caused by cracks and spacers, which again do depend on the particle type. However, it must be noted that such corrections are generally small for the ZEUS calorimeter, which makes the condensate energy a good first order approximation of the actually deposited energy. The particle type dependent improvements are applied to the reconstructed objects after the particle identification.

The features of the condensates found in phase 1 can be described as follows. In general terms, a condensate can correspond to the shower of a single particle, if it is isolated. Isolated electrons and photons are virtually always found as one condensate, due to the compact nature of their showers. Isolated hadrons are usually found as one condensate, but can also be split over two or more, due to the large fluctuations in hadron showers extending over many cells. Isolated muons can usually be reconstructed without problems, but due to their low energy deposit and the adjacency definition, they can extend over more than one condensate. This depends on their exact trajectory, muons not originating from the nominal interaction point, such as cosmics, are usually not reconstructed as one single condensate. Overlapping showers of nearby particles, or particles in dense jets, are usually merged into one condensate. This is caused by the simple geometrical way of clustering, which does not take the showering properties of different particles into account.

The expected features of condensates as described above, were confirmed in an analysis of the correspondence between generated particles and reconstructed condensates in Monte Carlo data. However, in the very forward region ($\theta < 10°$), where the particle density is high, many particles are usually merged into one condensate.

Condensates in Phase 2

In the reevaluation of calorimeter data, performed in phase 2 of the reconstruction, the condensate analysis is refined in order to better reconstruct single particle showers. In the analysis of 1992 data, this consisted only of the merging of incompletely reconstructed condensates, such as those with a single HAC cell, with other condensates. In the future, this procedure will be refined and extended in order

to separate, if possible, the various particles in a condensate, in those cases where their showers partially overlap.

Output Table

The condensates of phase 1 are stored in table CConSa as shown in figure 6.21. A relationship with the cell table Caltru defines the cell-condensate assignment. After the reevaluation of CAL data in phase 2 of the reconstruction program, the condensates are stored in table CuPaOb, with a similar structure.

6.9 Global Clustering and Jets

We use the term *global clustering* for a clustering of the calorimeter cells into, possibly large, jet-like objects, by applying a jet algorithm to the data. In this case, cells at a sizable physical distance of each other can still be merged into a single cluster, the merge criteria being aimed at clustering the fragmentation results of single partons.

The concept of a jet itself is not well defined; we will come back to this in the evaluation of the jet algorithms as presented in section 6.14. We will use an operational definition here: a jet is an object found by a jet algorithm; for a physical interpretation we will limit ourselves to the concept of a jet as a dynamical object that results from some mechanism that guides particles, emerging from collisions at high energy, into narrow cones.

The calorimeter is an important detector component in the measurement of jets, and, vice versa, jet reconstruction is an important part of the event reconstruction in the calorimeter. In the ideal case, the input to a jet finding algorithm should consist of the full set of 4-vectors of an event, such as available in Monte Carlo event generators. Although it is the task of an event reconstruction program to reconstruct those 4-vectors as closely as possible from detector information such as vertices, tracks, and clusters, the set can usually not be completely reconstructed in real data. In experiments at e^+e^- colliders, where events are relatively clean, jet finding algorithms are usually applied to fully reconstructed tracks. In experiments at pp or $p\bar{p}$ colliders, the input to the jet finder usually consists of energies in calorimeter cells. In phase 1 of the ZEUS calorimeter reconstruction program, we apply such a jet finding algorithm to the calorimeter cells only. Our starting point is formed by the calibrated cell energies from table Caltru. We do not make use of the condensates in order to prevent any biases that might be introduced by the cuts used in the condensate finding. We also do not make use of reconstructed tracks in the tracking detectors, as these were not fully operational in 1992 (see section 8.2). The momentum resolution of the tracking detectors is such that only tracks of charged particles of low or medium momentum are accurately reconstructed, the accuracy for high energy tracks lies below what can be reached with the calorimeter. Due to the influences of the magnetic field, charged particles at low energies may be

deflected far away from the jet axis. When the jet finding procedure uses only the calorimeter, these particles may not be assigned to the correct jet. The influence of this effect on the accuracy of the jet finding remains to be studied. However, at high jet energies, the calorimeter is more accurate than the tracking detectors, and it measures also neutral particles. The jet finding is performed globally in the whole calorimeter, i.e. also over the FCAL-BCAL and BCAL-RCAL boundaries.

The difference between local clustering into condensates and global clustering into jets is thus clear. The condensate algorithm only makes use of the local structure of the energy deposits by looking at energy sharing between neighbouring cells, and tries to reconstruct showers from single particles. The jet finding algorithm uses the global event topology in order to reconstruct the remnants of the partons that resulted from the hard scattering, and may combine many particles into one jet.

The transverse momentum of the hadronic remnants with respect to the jet axis is typically limited to a few hundred MeV. As a consequence, jets with energies below a few GeV have a large opening angle. They also have many particles with low energy. Jets from the interesting high Q^2 DIS events are typically well collimated and have a high energy. Taking into account the cross section, we expect at HERA many jets with low energy in the forward direction. The aim is to reconstruct the jets in the interesting high p_T events as well as possible, but also to perform well on the jets with low energy. We have implemented three kinds of jet finding algorithms:

- A fast phenomenological algorithm as used in the second level calorimeter trigger (**SLCT**);
- Several variants of the **JADE** algorithm, including the newly proposed k_T algorithm;
- A fixed cone algorithm in pseudorapidity (η)-azimuth (ϕ) coordinates, as commonly used in hadron collider experiments (**UA1, CDF**).

We will describe these algorithms below.

The energy of a cluster is determined by summing the energy deposit in the cells assigned to the cluster. For the cone algorithm, the position of the jet centre is taken from the calculated cluster centroids in (η, ϕ) space; for the other algorithms, the reconstructed position of the cluster centre is an energy weighted average over the centre of gravity of the cells assigned to the cluster. Where appropriate, the direction of the cluster in the calorimeter is determined from the cluster centres in EMC and HAC.

6.9.1 The SLCT Algorithm

The SLCT jet algorithm is an off-line implementation of the clustering algorithm used in the ZEUS second level calorimeter trigger. A full description is given in [69].

The algorithm is essentially a cone algorithm operating in polar angle (θ)-azimuth (ϕ) coordinates, with a cone size depending on the jet energy. The average

transverse momentum of the particles in the jet with respect to the jet axis is assumed to be 300 MeV, independent of the jet energy. A jet can then be modeled as a cone around the jet axis, whose opening angle decreases with increasing jet energy, since p_T in the jet is fixed, whereas p_L increases with the jet energy. The model assigns a canonical cone size (opening angle) α_c to a given cluster with energy E, following:

$$\alpha_c = \frac{a}{E} + b \tag{6.2}$$

where a is determined by the 300 MeV average transverse momentum in the jet, and a safety factor to ensure that virtually all energy of the jet is taken into account, and where b is essentially determined by the shower width of a single hadron.

These observations do not hold for electromagnetic showers. At HERA, neutral current DIS events are characterized by a scattered electron in the calorimeter. For this reason, the SLCT algorithm also tries to reconstruct single electromagnetic showers. These showers are recognized as clusters with more than 90% of the energy deposited in the EMC cells of the calorimeter. In these cases, the canonical cone size of the cluster is no longer given by equation 6.2, but is fixed, independent of cluster energy, to a value appropriate for electromagnetic showers. In addition, the implementation of the algorithm in the reconstruction program makes an exception for muon candidate clusters in the calorimeter.

The reconstruction efficiency and reliability has been studied with simulated DIS events with $Q^2 > 100$ GeV^2 in [69].

6.9.2 Variants of the JADE Algorithm

The original JADE algorithm comprises a division of the particles in an event into jets following a cluster criterion based upon the invariant mass between particle pairs [70]. In time, a number of modifications have been proposed to this scheme [71, 72, 73], without substantially modifying it, and therefore we shall refer to such algorithms as variants of the JADE algorithm. The LUCLUS routine in JETSET [74] offers a ready-to-use implementation of the JADE algorithm, and in addition, a similar algorithm based upon a distance criterion that is almost equivalent to transverse momentum between particles.

Our implementation of these algorithms operates on calorimeter cells, which are represented as momentum 4-vectors. The energy measured in the cells form the fourth component of these vectors, the momentum 3-vector that forms the spacial part points to the center of gravity of the cells. The mass of the 4-vectors is not given by the calorimetric energy measurement, and can be chosen to equal zero, or the charged pion mass. We denote such 4-vectors as $p = (\vec{p}, E)$. These 4-vectors are assigned to jets, based on the relative distance between the 4-vectors. The concept of distance needs a definition, or distance criterion, several of which are described below.

The jet finding procedure goes as follows. The relative distance between all possible pairs of particles (or cells) is determined. The pair with the smallest distance

is merged first into a *pseudoparticle.* This pseudoparticle now takes part in the procedure as if it were an original particle. The procedure of distance calculation between all possible pairs of particles (or pseudoparticles) and subsequent merging of pairs is repeated, until all remaining possible combinations have a distance larger than some cut-off. The resulting pseudoparticles are then the clusters.

We have implemented the distance criteria described below, including all mentioned options. For this purpose, the original LUCLUS code was modified, and interfaced to read from and write to ADAMO tables. The eventual results of the cluster finding appear (after cluster identification) in table **Cidclu**. The relationship to the individual cells in Caltru making up the cluster is stored in table Cidcel, as shown in figure 6.21.

The LUCLUS Algorithm

The LUCLUS distance criterion is:

$$d = 2|\vec{p}_1||\vec{p}_2|\sin(\theta_{12}/2)(|\vec{p}_1| + |\vec{p}_2|) \tag{6.3}$$

where θ_{12} is the angle between the two particles [75]. For small angles, this roughly corresponds to the relative transverse momentum between the pseudoparticles. Optionally, any time two pseudoparticles have been joined, cells can be reassigned to the pseudoparticle they are now closest to.

Invariant Mass based distances

In these kind of algorithms, the invariant mass squared between pairs of pseudoparticles is used as a distance:

$$d = (p_1 + p_2)^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2 \tag{6.4}$$

The original JADE algorithm calculates this distance neglecting the particle masses:

$$d = 2E_1E_2(1 - \cos\theta_{12}) \tag{6.5}$$

Several modifications to this scheme have been proposed, concerning either the calculation of the invariant mass with or without accounting for the particle masses, or the merging of two pseudoparticles (i.e. adding the 4-vectors). These are known as the E, E0, P and P0 schemes, whereby the E0 scheme is equivalent to the original JADE scheme [72].

It is convenient to scale the final distance cut-off with some scale that characterizes the event. For HERA, choices include the centre-of-mass energy s, but this is fixed for a given beam setting, the momentum transfer Q^2, the invariant mass of the hadronic system W^2, or the energy E_h of the hadronic system.

The k_T Algorithm

The most recently proposed change to the JADE scheme is known as the *Durham* or k_T algorithm [73]. Instead of using the invariant mass, the distance is calculated as:

$$d = 2\min(E_1^2, E_2^2)(1 - \cos\theta_{12}) \tag{6.6}$$

This criterion has emerged from the observation that the original JADE algorithm sometimes artificially introduces jets by clustering very soft gluons with very different directions. This effect is suppressed by the distance definition of equation 6.6.

6.9.3 The Cone Algorithm

Experiments at hadron colliders usually define jets as large energy deposits in cones of fixed radius in a metric based on a grid in pseudo rapidity ($\eta = -\ln(\tan(\theta/2))$) and azimuth ($\phi$) space. The cells of the ZEUS calorimeter do not correspond to fixed bins in η or ϕ. Instead, for each cell we take the actual cell centre coordinates, which implies that the bin size is not constant, but depends on the location in the calorimeter.

These kind of *cone* algorithms are used in several variants in various experiments [71]. Recently, one variant has been proposed as a standard in hadron collider physics [76]; this algorithm is a slight variation of the algorithm used by CDF [77]. This algorithm has been implemented for ZEUS in the following way:

1. The cells are ordered in decreasing transverse energy E_T, or optionally in energy E. Cells above a seed cut in E_T (E) can serve as *seed* cells.

2. A loop is performed over the seed cells, cells within a radius of $R_{precluster} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ of each other are grouped into preclusters. In this formula, $\Delta\eta$ and $\Delta\phi$ denote the distance between the cells in $\eta - \phi$ coordinates.

3. The centroid of each precluster is determined using the E_T (E) weighted $\eta - \phi$ centres of the cells assigned to it. For each precluster, a loop is performed over cells with E_T (E) above a low cut, these cells are assigned to the precluster if they are within distance R_{jet} (same definition as $R_{precluster}$) of the cluster centroid. The centroids of the clusters are then recalculated with the new cell list. This procedure is repeated until the cell-cluster assignment is stable. If stability is not reached, the procedure is stopped after a maximum number (20 by default) of iterations has been reached. However, in all tests that have been performed with this algorithm, this procedure has converged within 4 iterations.

4. Cells can be assigned to more than one cluster. These overlaps are resolved by calculating the amount of overlapping energy. If this is more than a fraction $f_{overlap}$ of the energy of the smallest cluster, the two clusters are merged,

otherwise the cells are assigned to the closest cluster. The centroids of the clusters are recalculated, and the resulting clusters form the jets.

6.10 Comparison between off-line and trigger clusters

A detailed comparison between clusters found in the off-line reconstruction and clusters found in the second level calorimeter trigger has been performed in [46] . We shortly list some of the differences.

1. The off-line reconstruction operates on UNO-calibrated (corrected) cell energies, whereas these calibration constants are not yet available on-line. As a result, cluster energies are more accurate.

2. The clustering in the trigger is severely constrained by a limited processing time. In order to attain this, higher cuts on cell and cluster energies are applied. As a result, less energy is clustered on-line, global energy sums and cluster energies are less accurate, and isolated muons (depositing only $\mathcal{O}(1)$ GeV in the calorimeter) are usually not found in the on-line clustering.

3. The position of the jet axis can be determined more precisely in the off-line reconstruction, since the full off-line calorimeter geometry is available. For this reason, a list of cell-cluster assignments is maintained in the off-line reconstruction. This has consequences in the accuracy of global sums like P_L^-.

In the evaluation of the 1992 performance of the trigger clustering, it was revealed that not all software was stable and bug–free [46]. The efficiency of electron finding has been shown to be slightly less than off-line. Although it is in principle possible to use clusters found in the trigger for reconstruction purposes, this has so far not been done for the reasons mentioned above. Whether this will be done at some time needs further study.

6.11 Cluster and Condensate Identification

The calorimeter reconstruction program makes an attempt to identify the found objects in terms of particles, or groups of particles that may have caused the shower. In section 4.3, we have described the different showering properties of various particles that can be measured in the calorimeter. The segmentation of the calorimeter, both longitudinal and lateral, can now be used to identify the reconstructed objects. In first instance, clusters and condensates will be labeled as belonging to one of the following classes:

- *'electron'.* This class consists in general of particles with pure electromagnetic showering properties, such as electrons, positrons and photons. The calorimeter by itself is unable to distinguish between them, but individual charged and neutral particles are easily separated using the tracking detectors.

- *'pion'.* This label will be given to showers that could be assigned to single, more or less isolated particles with hadronic showering properties, i.e. hadrons in general. The calorimeter cannot distinguish between the various hadrons that may enter the calorimeter, nor, by itself, between neutral and charged hadrons.

- *'muon'.* Muons in this context are defined as isolated tracks of minimum ionizing energy passing through the full depth of the calorimeter.

- *'jet'.* Any composite object that can not be attributed to single particles, but has jet-like properties and extends over a sufficiently large part of the calorimeter, is called 'jet'. These objects may be actual jets, but may also consist of a few randomly overlapping hadron showers.

- *'unknown'.* Incompletely reconstructed objects may not correspond to any known class; examples are single HAC1 or HAC2 cells above threshold. These get labeled 'unknown'.

For each cluster or condensate we calculate a number of 'shape parameters'. These shape parameters are chosen such that they quantitatively describe various properties of the reconstructed shower in terms of longitudinal and transverse shape. Such shapes are different for different types of particles: particles that shower purely electromagnetically, such as electrons and photons, have a short and narrow shower, hadrons tend to have a wide, irregular, and elongated shower, and muons do not shower at all. Therefore, these parameters distinguish between the various classes, and can be used as the discriminating variables by making appropriate cuts.

We list the shape parameters used in the program. Note that the list holds for condensates as well as for clusters, even though we generally denote the objects here as 'cluster'.

1. The total energy of the cluster and the energy per calorimeter section E, E_{emc}, E_{hac1} and E_{hac2} (if available). The magnitude of the deposited energy itself is only meaningful in the recognition of muons, which behave in the calorimeter as minimum ionizing particles. For other particles, energies are only used when scaled with other energies; this will be described below.

2. The total number of cells n_{tot}, and the number of cells of each calorimeter section, n_{emc}, n_{hac1} and n_{hac2}, contributing to the cluster.

3. The ratio E_{emc}/E between the energy deposited in the EMC cells of a cluster, and the total cluster energy.

4. The ratio E_{emc4}/E_{emc} between the energy deposited in the 4 most energetic EMC cells belonging to a cluster, and the total energy deposited in the EMC cells of that cluster.

5. The ratio E_{hac1}/E_{hac2} (if available) between the energy in a HAC1 section of a cluster, and the energy in a HAC2 section.

6. The *radius* of a cluster, calculated as the average distance of the cluster cells to the line connecting the reconstructed cluster centre and the vertex, is a measure of the shower width.

Electrons are recognized with the following cuts, more details are given in the next subsection on the efficiency of electron–pion separation: $E_{emc}/E > 0.85$, $E_{emc4}/E_{emc} > 0.85$, $n_{hac2} = 0$, $n_{hac1} \leq 4$, and $n_{emc} \leq 10$ if $E < 5$ GeV, $n_{emc} \leq 12$ if $5 < E < 40$ GeV, and $n_{emc} \leq 14$ if $E > 40$ GeV.

The recognition of muons is not one of the primary tasks of the calorimeter, but nevertheless quite interesting. The identification of muons is based on the typical track-like energy deposit of a muon penetrating the calorimeter. The total energy deposit for muons is restricted to lie between 0.7 GeV and 6 GeV. Limits on the number of EMC, HAC1 and HAC2 (if not in RCAL) cells are used, as well as the energy deposited in those sections. The ratio E_{hac1}/E_{hac2} characteristically peaks around 1 for minimum ionizing particles [2], and is therefore also used in the muon recognition in FCAL and BCAL. More details on muon–pion separation are given in subsection 6.11.2.

The separation of showers from single hadrons ('pions') from jets is simply based on the size of the reconstructed cluster, characterized by the number of cells (in particular EMC cells) contributing to the cluster, and the cluster radius. In order for a cluster to be identified as 'pion', it must not be recognized as electron or muon, consist of at least two calorimeter cells, have at most 6 EMC cells and at least 1 HAC1 cell, a $E_{emc}/E < 0.92$. A 'jet' is a cluster not identified as electron, muon or pion, with an energy exceeding 4 GeV, more than 7 cells, more than 4 EMC cells, and a radius of more than 15 cm.

If a cluster fails all mentioned cuts, it is labeled 'unknown'.

The information on the resulting classification is stored in the class attributes of the appropriate tables: **Cidclu** for the jets, and **CConSa** and **CuPaOb** for the condensates. The obtained identification results are used in the subsequent steps in the reconstruction, in particular the refining of the reconstruction of 'electrons'.

6.11.1 Multivariate Discriminators

One can try to combine the information contained in the shape parameters in order to find new and more effective discriminators. In particular, it is useful to think of

[2] The HAC2 cells in the outer modules in FCAL are smaller than the HAC1 cells. In these cells, we scale the deposited energy to the full cell size. If a HAC2 section does not exist (in RCAL), the ratio E_{hac1}/E_{hac2} can obviously not be used.

the N shape parameters as spanning an N-dimensional hyperspace, where distinct classes of clusters and condensates populate distinct regions. The problem of separation of the classes then reduces to finding the most optimum hyperplanes dividing the space in different regions, each region corresponding to a class. In particular, combining the information contained in the N parameters will generally be more effective than just applying cuts in each of the N parameters individually. Such problems have been known and studied for a long time. We briefly describe two methods that have been tried in this application.

Canonical Discriminant Analysis

In canonical discriminant analysis [78] one tries to find the optimum hyperplane that separates the classes, by making a linear combination of the shape parameters. The restriction of a linear combination implies that the hyperplane is flat, i.e. not curved. In this case, the coefficients of the linear combination can be calculated exactly, from the covariance matrices of appropriate *training samples*. Training samples contain representative samples of sufficient size of objects that belong to the classes to be separated, and for each object the class to which it belongs is known. The training samples thus serve as an example: they teach the method how to separate between the classes. In canonical discriminant analysis, the linear combination of the shape parameters forms a new variable, the *canonical variable*, that can be interpreted to represent a new axis in N-dimensional space, along which the classes are maximally separated, and which is perpendicular to the separation hyperplane. The calculation of the coefficients of the linear combination involves a maximization procedure of the distance between the projections of the classes on the new axis. When more than two classes are to be separated, additional uncorrelated canonical variables can be constructed. The canonical variables are reconstructed from the training samples only, before their actual use in a separation problem on real objects. An advantage of a linear method is the fact that the coefficients of the linear combination (the *canonical coefficients*), when appropriately normalized, are easily interpreted as representing the importance of a certain shape parameter in the discrimination, i.e. its discriminating power. This information could eventually be used in other applications.

We have tried the method in the separation of electron and pion induced showers in the calorimeter [79], with good results. We have found that the discrimination is almost entirely determined by only two of the shape parameters, which have subsequently been used in the implementation of electron-pion separation in the reconstruction program.

Neural Networks

A logical extension of the previous method is the use of a *neural network*. With the use of one or more hidden layers in the network, curved separation hyperplanes are possible, which may lead to a better discrimination. As in canonical discrim-

inant analysis, appropriately constructed training samples teach the network how to discriminate between the classes, before the actual use of the network in a separation problem of real objects. However, the coefficients that play a role in the network can not be calculated nor interpreted in a simple way. They result from a long, iterative learning procedure, in which the objects in the training samples are repeatedly fed to the network while forcing a particular output. In this procedure, the coefficients in the network are regularly upgraded, until they converge to their final values. There are many parameters that play a role in the learning procedure, and the coefficients may converge to a value that is not the optimum in the separation. We have tried a feed forward network with one hidden layer and back propagation training in the separation of electrons and pions, with results that are identical to the results obtained using canonical discriminant analysis. We conclude that the classes of electrons and pions in the calorimeter are well separated by a flat hyperplane.

6.11.2 Efficiency

The condensate and cluster identification performance has been tested with Monte Carlo generated single particles, with the full detector simulation turned on. Single electrons, pions and muons were generated, randomly distributed in θ and ϕ over the whole detector, except for the beam pipe holes. The particle energies were varied between 1 and 100 GeV (50 GeV for electrons). The obtained distributions of the 'shape parameters' have also determined the values of the mentioned cuts.

Electron-Pion Separation

The separation of electrons and pions is essentially determined by two of the shape parameters described above, namely by the ratios E_{emc}/E and E_{emc4}/E_{emc}. They describe the longitudinal and lateral shower development. The importance of these parameters in electron-pion separation was confirmed in a study using canonical discriminant analysis [79]: the correlation between these parameters and the resulting canonical variable is very high. In the separation, we use in addition to these two parameters, a limit to the number of EMC and HAC cells that may contribute to an electron cluster. Clusters or condensates are identified as 'electron' if they satisfy all cuts on the shape parameters.

The electron-pion separation performance is summarized in figure 6.6. By varying the cuts on the shape parameters, the electron recognition efficiency and the pion misidentification probability can be varied. A high electron identification efficiency implies loose cuts, and thus a relatively high pion misidentification probability, and vice versa. The separation performance is poorest at low particle energies, as expected, but improves quickly as the particle energy increases.

In the reconstruction program, the cuts are tuned for a 98% electron identification efficiency for electrons between 1 and 50 GeV (see previous section). Such a high efficiency needs relatively loose cuts, and therefore a non-negligible pion

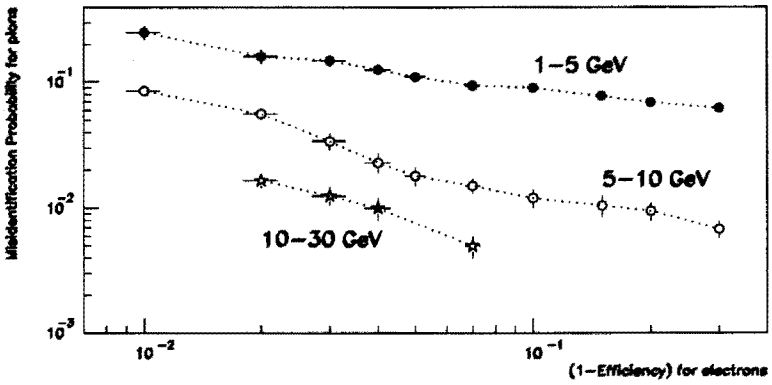

Figure 6.6: *The probability of misidentification of a pion as a function of the electron recognition efficiency, averaged over electrons and pions in various energy ranges.*

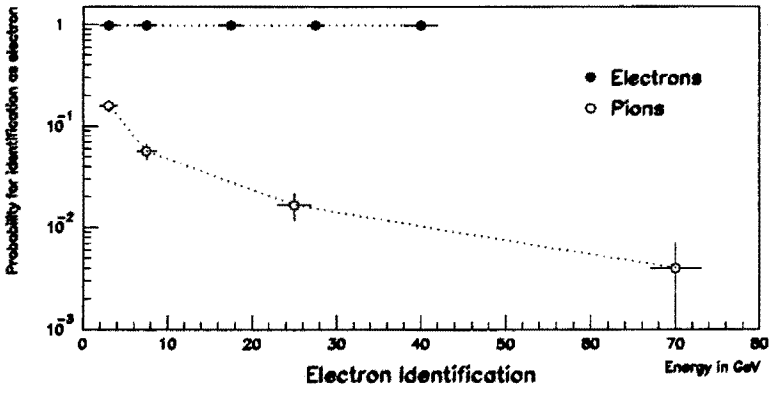

Figure 6.7: *The probability of identification of a cluster as an electron, as a function of cluster energy, for generated electrons and pions. The electron recognition efficiency is fixed to 98%.*

misidentification probability. This is shown in figure 6.7. The electron recognition efficiency is 98%, and the probability of misidentifying a pion as an electron decreases with increasing energy.

Muon-Pion Separation

Muons are easily separated from electrons and jets, but are less easily distinguished from showers caused by single hadrons of low energy. The muon recognition cuts have been described in the previous section.

For a total number of 5000 generated muons, 52% were uniquely identified as

'muon' based on the reconstructed condensates. However, a large fraction of the muons that were not recognized can be ascribed to an incomplete condensate reconstruction: if a muon traverses only a small corner of a cell, the deposited energy will be too small to pass the noise suppression cut, and the muon will be reconstructed as two separate condensates. Each of these two condensates will lack a HAC1 or a HAC2 cell, and thus be labeled 'unknown'. With the additional restriction that the condensate must have at least one HAC1 cell *and* at least one HAC2 cell, 85% of the muons were recognized. The probability of misidentification of a pion as a muon is only non-zero for pions between approximately 0.5 and 8 GeV, and amounts in that region to 1% of the generated pions. Note that this number is approximately equal to the probability of a pion, generated at the nominal vertex, decaying into a muon before reaching the calorimeter. The majority of the muons that were not recognized failed to pass the cut on the ratio E_{hac1}/E_{hac2} due to some large fluctuation in one of those energies. For a higher efficiency, this cut can be loosened. Using clusters, the muons are usually reconstructed completely, with a similar identification efficiency.

The cuts on parameters that use the HAC2 cells can not be used in RCAL. Since the other cuts are used without modification, this implies a higher muon recognition efficiency in RCAL, but also a higher probability of pion misidentification.

The reevaluation of CAL data in phase 2 of the reconstruction program will be used to refine the condensate reconstruction, in order to decrease the number of incompletely reconstructed condensates. This refinement will also decrease the fluctuations in E_{hac1}/E_{hac2} by a better reconstruction of the individual E_{hac1} and E_{hac2} values.

6.12 Reconstruction of showers of electrons and photons

The reconstruction of electrons is of crucial importance for HERA physics. The standard NC events will have a scattered electron in the final state; its position measurement is immediately reflected in the accuracy of reconstructed Q^2. In addition, many other interesting processes can have a final state electron or high energy (prompt) photon.

6.12.1 Additional reconstruction methods

The typical transverse size of an electromagnetic shower is related to the Molière radius. For the ZEUS calorimeter, this size is typically a little smaller than the size of the calorimeter cells. Therefore, a position reconstruction method based on an energy weighted average of the cell centers contributing to the cluster or condensate only, will give poor results. The method can be improved in the following way.

Each calorimeter cell is read out by two PMT's on opposite sides. The light created locally in the scintillator strips by a traversing particle, is attenuated in the

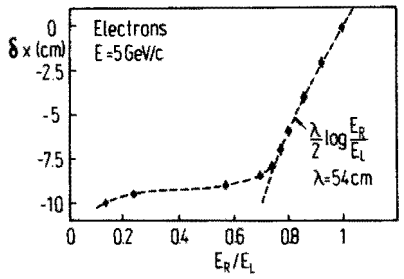

Figure 6.8: *Distance δx between the centre of gravity of the deposited energy and the middle of an EMC section, as a function of the ratio E_R/E_L of right and left readout energies.*

scintillator material while traveling towards the wavelength shifter bars on the side of the modules. The ratio E_R/E_L of energy measured on the right [3] side of a cell and energy measured on the left side of a cell, therefore depends on the location of the energy deposit. This can be used in a refinement of the reconstruction of the cluster position in one direction in the calorimeter (x in FCAL and RCAL, ϕ in BCAL).

Figure 6.8 shows test beam results for this ratio E_R/E_L as a function of the impact point of an electron beam, for the ZEUS prototype calorimeter [48]. Assuming an exponential light attenuation, the distance δx between the centre of gravity of the deposited energy and the middle of an EMC section can be calculated according to

$$\delta x = \frac{\lambda}{2} \ln \frac{E_R}{E_L} \tag{6.7}$$

where λ represents the scintillator attenuation length. Figure 6.8 shows that, in the prototype, δx is a logarithmic function of E_R/E_L with $\lambda = 54$ cm, except close to the boundaries of the cell, where additional effects such as the generation of Čerenkov light by a shower particle travelling through the wavelength shifter disturbs the logarithmic behaviour. Similar studies have been performed for the real calorimeter modules in ZEUS (with lead sheets between the modules), and the resulting correction δx is used in a refinement of the position reconstruction for electrons.

In the other direction (y in FCAL and RCAL, z in BCAL), the calorimeter cells are smaller (5 cm in FCAL and BCAL, 10 cm in RCAL). Showering electrons deposit energy in more than one cell in this direction, even though the central cell, where the electron enters, still contains the major part of the energy. The ratio E_{above}/E_{below} of energy in the cells immediately above and below the central cell,

[3] We denote the module sides here as 'right' and 'left', these correspond to the terminology 'high' and 'low' of section 6.2, but will only be used in cases where the actual meaning of 'right' and 'left' does not matter.

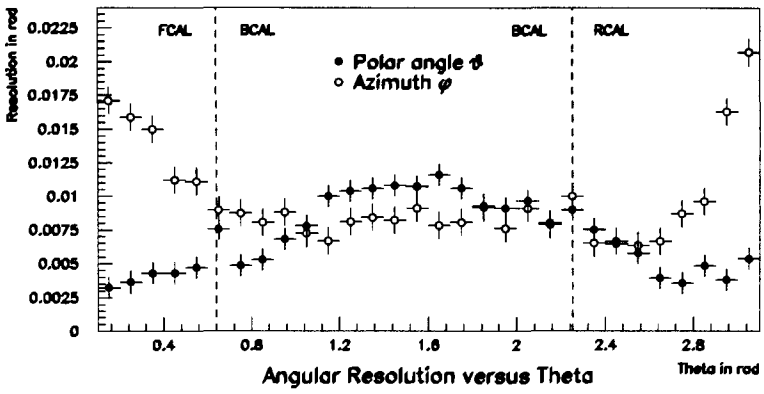

Figure 6.9: *Angular resolutions for the electron position reconstruction as a function of polar angle θ*

is a good measure of the centre of gravity of the shower, and is thus used for a refinement of the position reconstruction in this direction.

6.12.2 Position Reconstruction Accuracy

Using the methods described, we have evaluated the position reconstruction accuracy with Monte Carlo generated electrons, randomly distributed in θ and ϕ over the whole detector, except for the beam pipe holes.

For electrons between 25 and 30 GeV, we have investigated the distributions of $x_r - x_t$ and $y_r - y_t$, where x_r and y_r represent the reconstructed values of x and y of the impact point of the electron in the calorimeter, and x_t and y_t represent the Monte Carlo true values of these quantities. The obtained width (standard deviation) in the x direction in FCAL was 1.2 cm, whereas in RCAL this was 0.8 cm. In the y direction, the width of the distribution in FCAL equaled 1.7 cm and in RCAL 1.2 cm. In the BCAL, the width of the distribution in z, which coincides with the direction in which the cells are smallest, equals 1.2 cm. These resolutions improve with increasing particle energy. We observe that the reconstruction accuracy is best in the direction where the difference between left and right energy can be used, even though the cells in the other direction are smaller in size.

For physics, it is more relevant to plot the resolutions in the reconstruction of the polar angle θ and the azimuth ϕ. These resolutions are plotted for electrons between 25 and 30 GeV as a function of θ in figure 6.9, and averaged over FCAL, BCAL and RCAL as a function of energy in figure 6.10.

In FCAL and RCAL, the polar angle θ is reconstructed better than the azimuth ϕ, due to the fact that a given inaccuracy in the reconstruction of x and y has a larger influence on ϕ than on θ, especially around the beam pipe. For BCAL, the situation

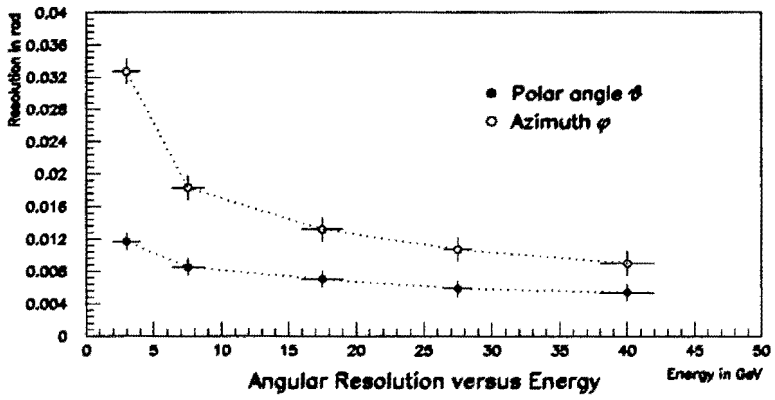

Figure 6.10: *Angular resolutions for the electron position reconstruction as a function of energy.*

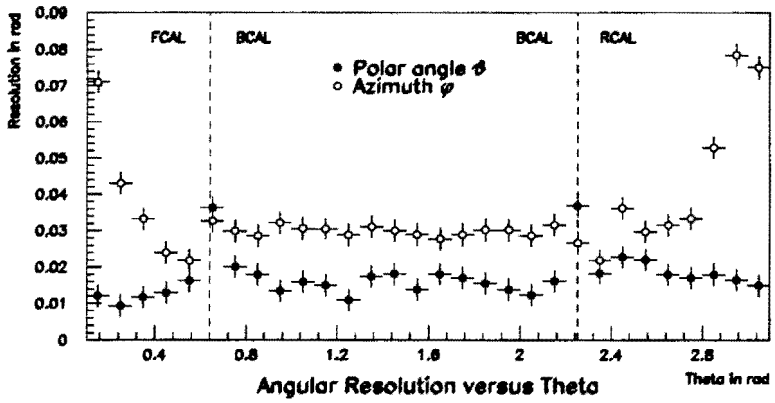

Figure 6.11: *Angular resolutions for the pion position reconstruction as a function of polar angle θ*

is reversed, due to the fact that the left/right energy difference contributes to the reconstruction of ϕ, whereas θ coincides with the z-direction. The resolution in θ is better than 5 mrad in FCAL and RCAL, and 10 mrad in BCAL. No systematic biases have been observed. At lower electron energies, the resolution degrades.

6.13 Reconstruction of Pion Showers

The position reconstruction resolution for pions as a function of polar angle θ and energy E is shown in figures 6.11 and 6.12. The obtained resolution is somewhat worse than for electrons. No use has been made of the difference in left and right

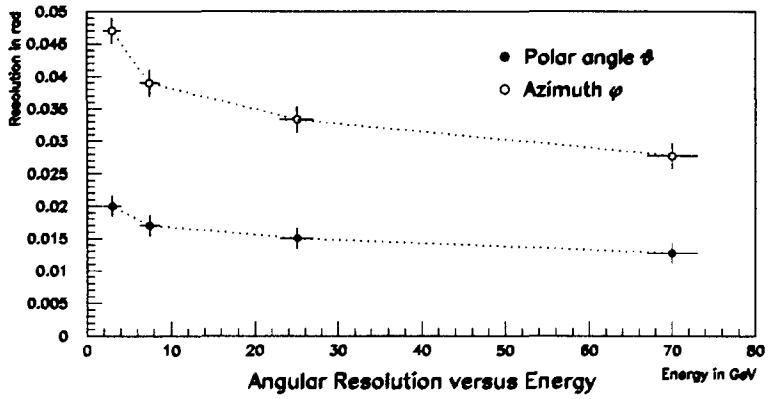

Figure 6.12: *Angular resolutions for the pion position reconstruction as a function of energy.*

energies of the cells contributing to the shower. Pion showers, on the other hand, occupy on average many more cells than electron showers, which has a positive effect on the accuracy of the position reconstruction. We observe that the resolution in θ for pions is degraded at the FCAL-BCAL and BCAL-RCAL boundaries, located at 0.64 rad and 2.25 rad respectively. These regions are characterized by a relatively high amount of inactive (support) material in front of the calorimeter, as shown in figure 4.12, and a gap between the various calorimeters themselves. For electrons, this effect is much smaller.

6.14 Efficiency of Jet Reconstruction

In this section, we evaluate the performance of the jet finders using Monte Carlo generated deep inelastic scattering events, with full detector simulation.

It is not immediately evident how the efficiency of jet finders should be judged. The concept of a jet itself is not well defined, but exists only within the context of a certain particle combination algorithm, or a measure of distance between particles. Jets carry the information of the partons involved in the hard scattering process, but the direct coupling between partons and jets as seen in the hadronic final state is obscured by initial and final state QCD radiation of the partons involved, by the hadronization process, and by detector effects. Within some of the current fragmentation models, it is not always possible to assign a produced hadron uniquely to any of the initial partons. In the LUND string model, for example, particles originate from strings between partons. Subsequently, it is not always possible to uniquely link a found jet with any of the partons.

We have chosen to solve this problem by performing jet finding on the parton level on a similar basis as on calorimeter cell level. The Monte Carlo generator

programs that have been used will be described in chapter 7. The Monte Carlo data we have used here consists of samples of events generated with LEPTO 6.1 (matrix element plus parton shower), and HERACLES+ARIADNE. For the application of comparing parton jets and calorimeter jets, no differences between the generators were observed, so that we can use both samples. The MRSD0 structure function was used, and cuts in the kinematical variables of $Q^2 > 10$ GeV^2 and $x > 10^{-4}$ were applied. In order to increase the statistics of high energy jets, an additional sample with $Q^2 > 100$ GeV^2 and $x > 10^{-3}$ was used in the evaluation of the cone algorithm.

The output of the Monte Carlo generator, as stored in table FMCKin of figure 5.2, consists of the original beam particles and the partons participating in the hard scattering, the partons in the final state after the QCD cascade has been simulated, either by parton showers or by colour dipole radiation, and the long lived particles that result from the fragmentation and decay of the final state partons.

We have verified that the collection of simulated final state partons after the QCD cascade conserves energy and momentum of the beam particles, and can thus be regarded as the 'true' final state on parton level. Note, however, that many of these partons are gluons of which it is impossible to determine from which original parton, going in or coming out of the hard scattering, they were radiated. We apply jet finding algorithms to this Monte Carlo partonic final state and cluster these into parton jets. We do the same for the calorimeter cells, and by using identical algorithms (cone, JADE), and identical parameters (cuts, cone sizes), we obtain jets that can be immediately compared to the parton jets. This comparison thus focusses on the effects of hadronization and on detector effects in the jet reconstruction.

In the evaluation of the JADE algorithm, we have used the distance measure of invariant mass, with a cut-off of $m_{ij}^2 \geq y_{cut} = 0.02\ W^2$. In this procedure, W^2 was calculated with the 'double angle formula' of equations 2.36, 2.37, 2.40 and 2.7. When the JADE algorithm is used for jet counting purposes, it is customary to vary the cut-off y_{cut}. However, we are only interested in a description of the final state in terms of jets, and the used cut-off was found to work well for HERA [80]. We have verified that variations in this cut-off between $0.005\ W^2$ and $0.05\ W^2$ have influences in the jet multiplicity, but not in the conclusions on energy and angular resolution.

Due to the fact that at HERA a large part of the longitudinal momentum disappears into the forward beampipe and is not measured, it has been found beneficial to include an extra pseudoparticle in the final state before the jet finding is applied [81]. This *beam pseudoparticle* compensates for the missing longitudinal momentum in the event, and will thus be aligned with the proton beam. This procedure ensures that also jets in the very forward area can be reconstructed well, and as a consequence, every event has at least 1 jet: the proton remnant jet. It is customary to denote a HERA event with N jets outside the target remnant area as a $N+1$ jet event, when the JADE algorithm is used.

It should be noted that the JADE algorithm is applied here to calorimeter

cells, and not, as customary, to reconstructed 4-vectors. This improper use may negatively influence the performance of the algorithm.

In the evaluation of the cone algorithm, a minimum E_T of 300 MeV was required in a cell in order to be considered as a cluster seed, and the jet cone radius R was fixed to 1 unit in $\eta - \phi$ space. This value for the radius was found to be the optimum for the energy and angular resolution of jet finding at HERA: when a smaller radius is used, too much energy is not contained in the cone, but when a larger radius is used, the energy and the position of the jet are affected by energy deposits far away from the jet axis. A low cut in the minimum E_T for seed cells introduces many preclusters, most of which are merged with each other in the jet finding. To some extent, therefore, the jet finding is not very sensitive to this cut. However, a too low value for this cut may lead to additional clusters not related to actual jets, whereas a too high cut may lead to actual jets not being found.

The cells assigned to the reconstructed scattered electron were removed from the cell list used in jet finding, so that the reconstructed jets did not include the electron. Since the cone and the JADE algorithms do not make any special precautions for electron or muon candidate clusters, in contrast to the SLCT algorithm, but treat all cells as part of a jet, and since the scattered electron is not part of the reconstructed clusters, we treat all clusters as jet candidates in this analysis. The minimum energy for a candidate to be accepted as an actual set was set to 4 GeV. The $\eta - \phi$ space used in the cone algorithm contains a singularity for $\theta \rightarrow 0$ and $\theta \rightarrow \pi$, i.e. in both beam holes. This is most severe for the forward beam hole, where the target remnant jet resides. The cone algorithm is not able to reconstruct a single target remnant jet. However, the jets that are reconstructed in the forward area are not meaningless: since the cone algorithm focusses on high transverse energy deposits, the reconstructed jets do correspond with actual particles, and only the definition of a 'jet' becomes meaningless. In this evaluation, however, we will circumvent these problems by explicitly excluding the very forward region $\eta > 2$, or $\theta < 0.25$ rad, from the jet finding. The value of this cut was motivated by the fact that, with a cone radius $R = 1$, the outer edge of the cone stays sufficiently away from the beam pipe. Such a cut is not necessary for the JADE algorithm.

In figure 6.13 we show for the cone algorithm and for the JADE algorithm, the distribution of reconstructed jets on parton level (open histograms) and calorimeter cell level (shaded histograms) in energy E and polar angle θ. The plots are normalized to an identical number of events. Generally, the jets on cell level follow the parton jets quite well, although the number of reconstructed jets on cell level is lower than on parton level. By including the extra missing-momentum-pseudoparticle, the JADE algorithm always explicitly reconstructs the target remnant jet. This jet is characterized by a very small, or zero, polar angle, and a very high energy exceeding 400 GeV. Such a jet is not reconstructed with the cone algorithm. The JADE algorithm always clusters all cells in an event, which leads to jets of possibly large

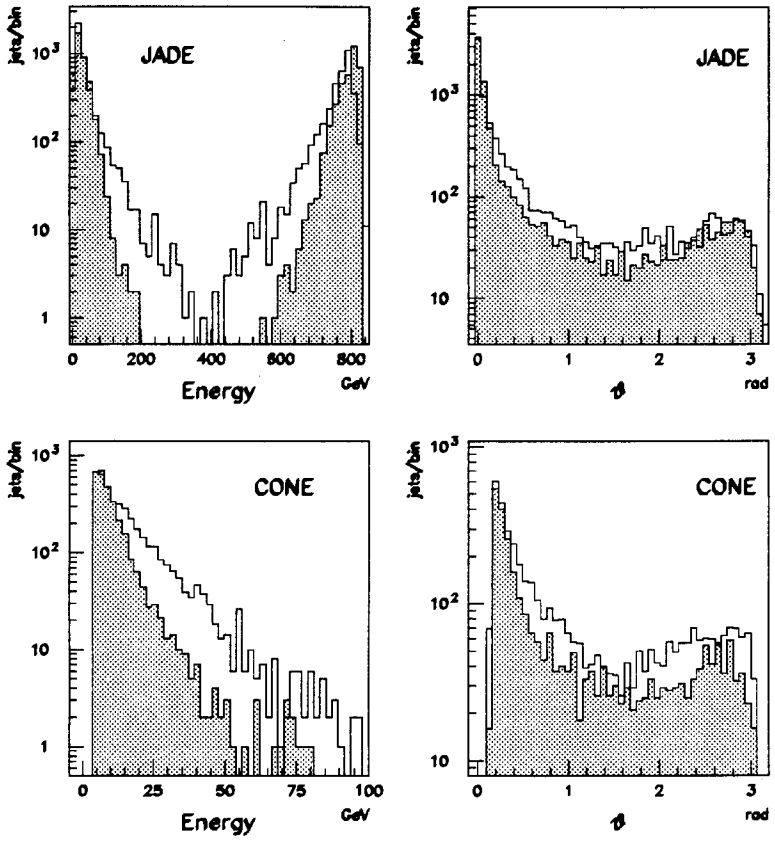

Figure 6.13: *Reconstructed jets on parton level (open histograms) and calorimeter cell level (shaded histograms) for the cone algorithm and the JADE algorithm, as a function of energy and polar angle.*

sizes and high energies, up to 200–300 GeV even for the non target remnant jets. In contrast, the cone algorithm does not look beyond the cone of predefined size, and is, due to the use of pseudorapidity, particularly sensitive in the forward region, where the jet energies are highest. Therefore, the cone algorithm reconstructs in general jets of lower energy. Also, not all cells need to be part of a jet. More details of the algorithms will be investigated in the next sections.

Angular Resolution

We define the angular resolution of the jet reconstruction as the width of the distribution of the difference in polar angle θ between the parton jet and the corresponding cell jet. In order to obtain a meaningful correspondence between parton jets

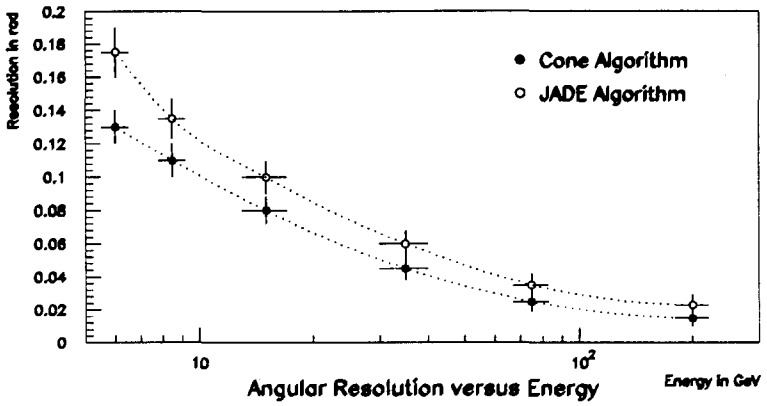

Figure 6.14: *Angular resolution of the jet reconstruction as a function of energy.*

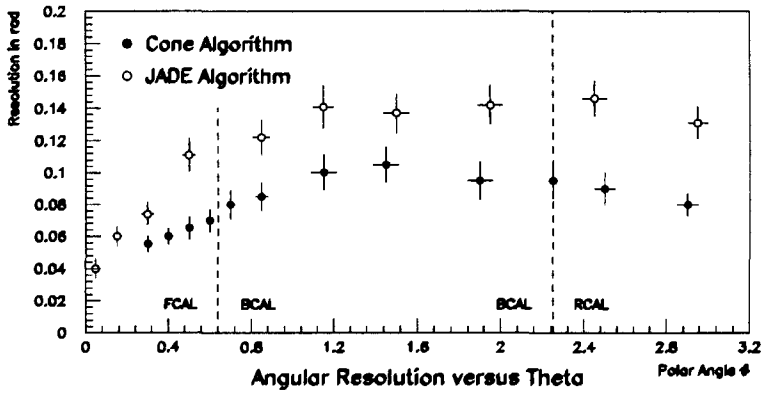

Figure 6.15: *Angular resolution of the jet reconstruction as a function of polar angle θ.*

and cell jets, we select one cell jet per event, and match this cell jet with the parton jet that has the smallest angular difference with the cell jet. In order to select not just any jet, but a relevant jet like the current jet candidate, the selected jet is chosen by a maximum value for the quantity $E \sin\theta$ of the jet, explicitly excluding the target remnant jet (all jets at $\eta > 2$ for the cone algorithm). In figures 6.14 and 6.15 the so obtained angular resolution is plotted for the cone algorithm and for the JADE algorithm as a function of jet energy and jet angle. Figure 6.14 integrates over jets at all angles, as they occur in the Monte Carlo data sample of $Q^2 > 10$ GeV^2. Similarly, figure 6.15 integrates over jets at all energies. The angular resolution improves with energy, as expected, to a value of better than 20 mrad for the cone algorithm at high energies. The JADE algorithm appears to be more sensitive to hadronization and detector effects than the cone algorithm, since the JADE al-

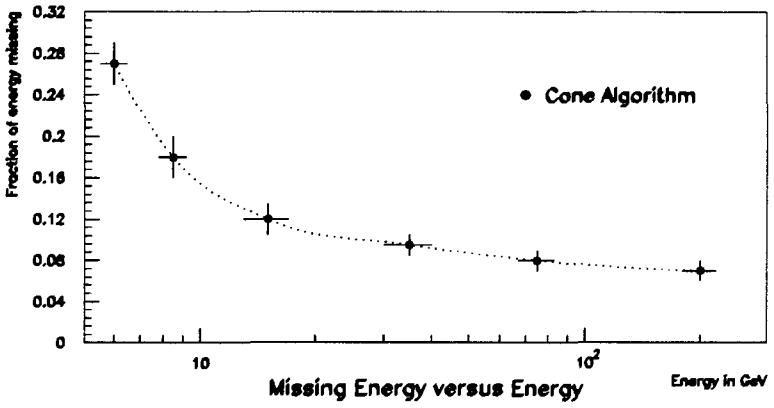

Figure 6.16: *Fraction of parton jet energy missing from the reconstructed jet as a function of energy.*

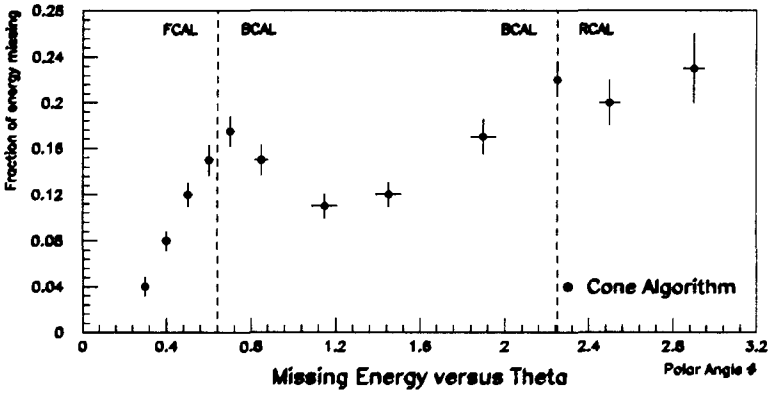

Figure 6.17: *Fraction of parton jet energy missing from the reconstructed jet as a function of polar angle θ.*

gorithm includes all cells in jets, even far away from the jet axis, whereas the cone algorithm makes a cut-off at the edge of the cone. The angular resolution is best in FCAL, note that the average jet energy is highest there.

Jet Energy

The energy of jets reconstructed on cell level is generally lower than the energy of parton jets. The fraction of energy missing, $(E_{parton} - E_{jet})/E_{parton}$, is a function of jet energy and jet angle, as shown in figures 6.16 and 6.17 for the cone algorithm. The fact that energy is missing from the jet can be explained by the observation that the jets have many particles of low energy, which loose a relatively large amount of

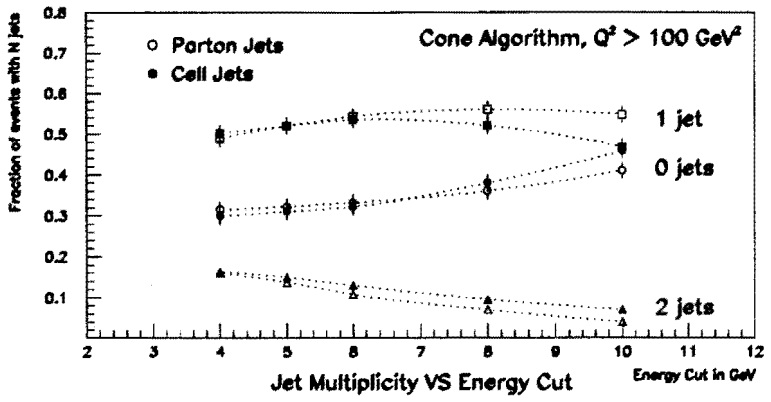

Figure 6.18: *Fraction of events with 0, 1 or 2 jets as a function of the minimum jet energy, for the cone algorithm, applied to events with* $Q^2 > 100\ GeV^2$.

their energy traversing inactive material in front of the calorimeter, and, if they are charged, may be bent far away from the jet axis by the magnetic field. As expected, the fraction of energy missing decreases with increasing jet energy. Again, the fact that on average less energy is lost in FCAL can be attributed to the fact that jets have generally more energy there. In addition, the influence of the dead material in front of the calorimeter can be observed, comparing figure 6.17 with figure 4.12.

The JADE algorithm (not plotted) shows larger and more irregular fluctuations when the cell jet energy is compared to the parton jet energy, confirming the conclusion that the JADE algorithm is more sensitive to fragmentation and detector effects, and that the overlap between parton jets and cell jets is worse.

The jet energy resolution, defined as the width of the distribution of $(E_{parton} - E_{jet})/E_{parton}$, has also been evaluated for the cone algorithm. The resolution gets better with increasing jet energy, improving from $\sigma/E = 0.24$ for jets between 5 and 7 GeV to $\sigma/E = 0.05$ for jets above 100 GeV.

Jet Multiplicity

We have investigated the number of reconstructed jets per event, on parton level and on cell level, as a function of the minimum energy required for a jet in order to be accepted. We plot the jet multiplicity as the fraction of events with N reconstructed jets of sufficient energy, where $N = 0, 1, 2$. In the evaluation of the JADE algorithm, we include all jets. For jets found with the cone algorithm, we demand $\eta < 2$.

Figure 6.18 shows the results for the cone algorithm, applied to events with $Q^2 > 100\ \text{GeV}^2$. It is clear that the number of reconstructed jets on cell level is approximately equal to the number of jets on parton level, except at high energy cut-offs. We have verified that this correspondence remains good when the cut-off on η is varied.

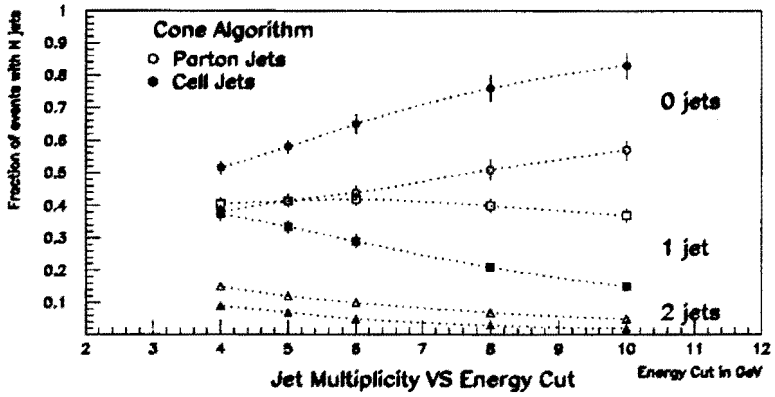

Figure 6.19: *Fraction of events with 0, 1 or 2 jets as a function of the minimum jet energy, for the cone algorithm, applied to events with* $Q^2 > 10\ GeV^2$.

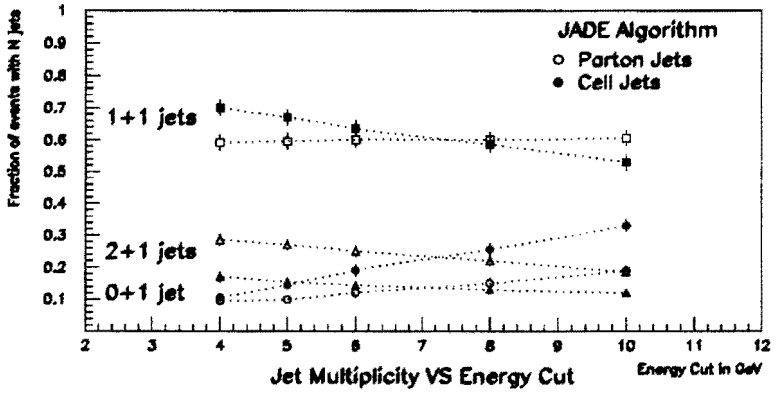

Figure 6.20: *Fraction of events with 0+1, 1+1 or 2+1 jets as a function of the minimum jet energy, for the JADE algorithm, applied to events with* $Q^2 > 10\ GeV^2$.

For events at lower Q^2, which are most abundant at HERA, the differences are larger. This is shown in figure 6.19 for the cone algorithm, and in figure 6.20 for the JADE algorithm. The fact that for the cone algorithm E_T is the important parameter, whereas for this Monte Carlo sample the average E_T is not very high, in addition to the cut $\eta < 2$, leads to the observation that most events have no jets with this algorithm. The jet multiplicity differs between parton and cell level, and is therefore sensitive to fragmentation and detector effects. The JADE algorithm always gives at least 1 jet, and usually more. The jet multiplicity with the JADE algorithm differs to a lesser extent between parton and cell level.

6.15 Finding the scattered electron

A common physics task is the recognition of the scattered electron in neutral current DIS events. In this thesis, which only covers the 1992 data taking period, we have used two complementary algorithms, both of which use the calorimeter only.

1. based on the reconstructed clusters and condensates. Clusters and condensates identified as *electron*, with an energy of more than 4 GeV are considered to be candidate for a scattered electron. If there are more candidates, the routine will prefer RCAL over BCAL over FCAL, and will take the one with the highest energy in that section.

2. based on the topology of the energy deposited in the cells of the calorimeter. Electron candidates are found as EMC cells with a high energy content. Around electron candidate cells, an inner and an outer cone are formed. The candidate is identified as electron based on the energy distribution in the cones, and in the EMC and HAC sections. The candidate with highest p_T is regarded as the scattered electron.

An evaluation of the algorithms has shown [82] that both have comparable performances.

6.16 Output tables

Figure 6.21 shows an ERD with the tables created by the calorimeter reconstruction program.

The recalibrated cell energies are stored in Caltru, which plays a central role in the reconstruction program. Instead of storing the energies of the 2 PMT's of a cell separately, we store the total energy in the cell, and the *imbalance*, representing the difference between the two channels. In addition, the reconstructed time of both PMT's is stored.

For the condensates, we store in CConSa the total energy E and the energy in the EMC, HAC1 and HAC2 sections, the reconstructed position of the centre of gravity of the condensate x, y and z (this can trivially be converted to θ and ϕ) and the individual reconstructed centres in EMC, HAC1 and HAC2, the result of the condensate identification procedure (the *class*), the direction cosines of the line connecting the condensate centre in the EMC with the condensate centre in the HAC section, the condensate radius (a measure of the lateral width), and the number of contributing EMC, HAC1 and HAC2 cells.

The phase 2 condensate table CuPaOb and the cluster table Cidclu have an identical structure. The relationship to the individual cells in Caltru making up the cluster is stored in table Cidcel.

Table CR1obj contains event related quantities like the total measured energy in the calorimeter, the scalar E_T, p_T (where cells are added vectorially), the number of

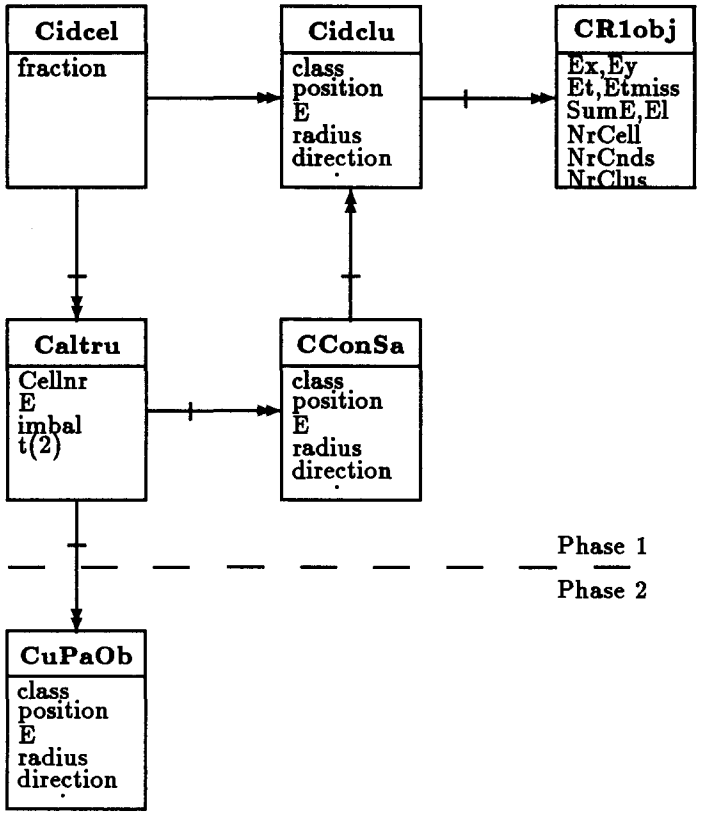

Figure 6.21: *The ADAMO Output of the Calorimeter Reconstruction*

cells with energy above the noise suppression cut, and the number of reconstructed clusters and condensates.

The different philosophy of condensates and clusters has been explained in section 6.9. The reconstruction program offers both kinds of objects to the physics analysis: condensates for an analysis whereby specific single particles in the calorimeter, such as muons, prompt photons, the scattered electron, or electrons from semileptonic decays of heavy quarks, are needed, and clusters for a jet analysis of the final state.

Chapter 7

Event Simulation

In this chapter we will describe some of the important points for event simulation at HERA, and the implementation of these points in a number of event generator programs.

Simulations of interactions between elementary particles and processes that play a role in the measurement of these interactions in a detector are usually performed with Monte Carlo techniques, and the resulting programs are known as Monte Carlo programs. In modern high energy physics experiments, such programs are necessary in the data analysis in order to estimate efficiencies, acceptances and biases. When one tries to measure processes that involve strongly interacting particles, the effects of the hard scattering are obscured by fragmentation and hadronization effects. Monte Carlo programs can help in unfolding these effects, in order to enable a comparison of the measured interactions with the underlying perturbative theory on parton level, such as QCD. As such there is also an interest in the models themselves that are used for the description of hard and soft gluon radiation and fragmentation. The processes that play a role there balance on the edge of what can be described perturbatively in QCD, and what can no longer be calculated perturbatively. It is therefore possible to speak of 'QCD-inspired models'.

Figure 7.1 gives a schematic representation of Monte Carlo event generation in ep physics [83]. The central part of the event generation consists of the simulation of the hard scattering process, where partons p_i and p_j scatter and produce partons p_k and p_l. This part can be calculated perturbatively. For ep scattering, examples of such processes have been described in chapter 2 of this thesis. The partons p_i and p_j involved in the hard subprocess are taken from the original electron or proton, which involves structure functions $f_{i/e}$ and $f_{j/p}$. After the hard scattering, the partons p_k and p_l, as well as the electron and proton remnants, must fragment and decay in order to produce the stable final state particles. In DIS, the original particle that emerges from the electron line is a γ, a Z^0, or a W; at low Q^2 the photon can behave as a hadron or as a $q\bar{q}$ pair and produce a remnant.

It has been shown in e^+e^- collisions [84] and in fixed target deep inelastic

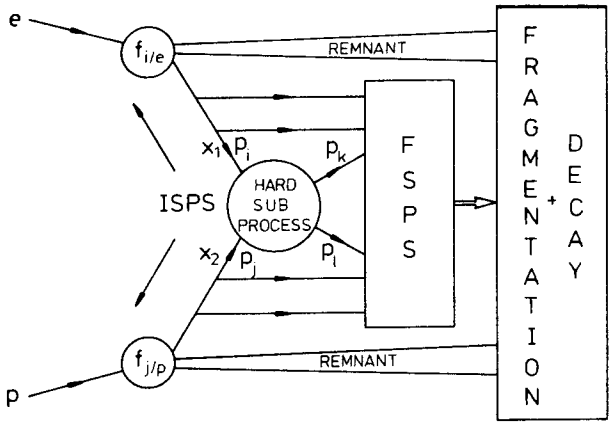

Figure 7.1: ***A representation of the generation of ep events.***

scattering [85, 86, 87], that a perturbative description of the hard scattering plus the use of a fragmentation model alone are not sufficient to describe the hadronic final state. In addition, the effects of multiple gluon emission, or QCD cascades, in the initial and final state have a measurable influence on the hadronic final state. One can try to calculate such corrections exactly, order by order in α_s, by calculating the exact matrix elements (ME). Such corrections are difficult, and only partial second order results exist for leptoproduction [88]. Although matrix elements are exact, their practical usefulness is limited by the lack of such higher order calculations. In deep inelastic scattering, the matrix elements for hard gluon radiation and boson-gluon fusion are known; they describe the first order (in α_s) corrections to deep inelastic scattering, and are therefore usually denoted as the first order matrix elements. In this thesis, we will use this convention, and denote the standard deep inelastic scattering process with a single struck quark leading to a single current jet as the 'lowest order' process, and processes that are first order in α_s as 'higher order' processes.

Instead of exact calculations using matrix elements, inclusion of higher order effects can be approximated by various models of such QCD cascades. The popular parton shower model [86, 89] is based on a summation of the leading logarithms in all orders. Figure 7.1 shows such parton showers: an initial state parton shower (ISPS) before the hard scattering, and a final state parton shower (FSPS) after the hard scattering. In the next section we describe two complementary models for QCD cascades that have been used in this thesis.

7.1 QCD Cascades

In e^+e^- experiments, the necessity of adding models of QCD cascades to the event generator in order to describe hadronic final states, has clearly been demonstrated [84]. However, in these experiments strongly interacting particles are only present in the final state, i.e. after the hard subprocess.

In final state parton showers, partons with large time-like virtuality produce showers of partons with lower virtuality, until some lower limit on the virtuality is reached and the partons hadronize. The evolution of initial state parton showers is considerably more complicated and cannot be tested in e^+e^- data. An incoming parton with low space-like virtuality radiates time-like partons, increasing its space-like virtuality in each branching, until it is struck by the electroweak boson in the hard scattering.

The amount of parton emission depends on the maximum virtuality Q^2_{max} of the showering parton just before and after the boson vertex. In a Monte Carlo simulation, these virtualities are chosen between zero and a maximum that is determined by the energy or momentum transfer scale in the process. In deep inelastic scattering, the scale for the partonic resolution is set by Q^2. However, if this were used as the scale of the cascade, processes at low Q^2, such as boson-gluon fusion, would show little or no QCD radiation. In such processes, the scale is related rather to the invariant mass of the produced quark-antiquark pair. One could therefore also take the hadronic invariant mass W^2 as the scale of the parton shower. However, since

$$W^2 = Q^2 \frac{1-x}{x} = ys - Q^2 \tag{7.1}$$

(neglecting the proton mass), W^2 is generally large at HERA when Q^2 is small (at low x). This implies that the amount of parton emission, and therefore the hadronic final state, should differ substantially between taking W^2 or Q^2 as scale. However, the use of W^2 is likely to overestimate the amount of parton radiation, since at low x, the major part $(1-x)$ of the energy in the final state is carried by the proton remnant, and does not contribute to the QCD radiation. Note that such an ambiguity in the choice of scale does not exist for e^+e^- physics, where $Q^2 = W^2 = s$.

In the parton shower model, special care must be taken to simulate coherence effects in the gluon emission. The separation into an ISPS and a FSPS neglects the effects of interference between the two. With parton showers, processes up to arbitrarily high orders in α_s are included, but only in the leading log approximation. The use of leading log approximations implies that the emission of soft gluons or gluons close to the direction of the emitting parton is well described, but that the emission of hard gluons at large angles could be mistreated.

An alternative approach to QCD cascades is given by the colour dipole model (CDM) [90]. In DIS, a colour dipole is formed between the scattered quark and the remnant diquark, which acts as a colour antenna and emits partons, in analogy

to electromagnetic radiation from a conventional antenna. Since the emitted gluon carries colour, the emission of a second softer gluon can be treated as radiation from two independent dipoles, one between the quark and the gluon and one between the diquark and the gluon. This is generalized so that further emission is given by more, independent dipoles. This formalism ensures that an ISPS arises naturally from the treatment of recoils in emitting dipoles and needs not be added explicitly. When a dipole radiates, it is not possible to decide which end radiated, and the process can be seen as the coherent sum of different Feynman diagrams. A problem arises in DIS where one end of the initial dipole (the proton remnant) is not point-like, but an extended object. This extended source leads to a suppression of gluon radiation, as radiation with large p_T, or small wavelength λ, sees only a fraction a of the energy in the extended source, and the rest of the source does not take part in the emission. In the colour dipole model, this fraction a is implemented as

$$a = (\frac{\mu}{p_T})^{\alpha} \tag{7.2}$$

where μ is a scale related to the inverse of the size of a hadron, and α a dimension that describes the energy distribution in the source, e.g. $\alpha = 1$ indicates a linear distribution.

7.2 Fragmentation

The fragmentation and hadronization of produced partons involves processes on a very low virtuality scale. Since α_s is large in that region, the processes can not be calculated perturbatively. Instead, several models exist that simulate the fragmentation. We will limit ourselves here to the LUND string model.

The LUND String Model

The LUND string model is implemented in the program JETSET [74]. The colour topology of the produced partons is reflected in the way they are connected by strings. A string connects a colour triplet (quark or anti-diquark) with a colour anti-triplet (anti-quark, diquark), via a number of gluons that can be seen as kinks on the string. The string therefore represents the colour flow between the partons. It is the string that fragments into hadrons, not the individual partons.

There are many phenomenological parameters that can be varied in this model. Strong correlations exist between the parameters that control the QCD cascade and those that steer the fragmentation, since the fragmentation can effectively compensate for what the perturbative phase fails to describe. Many parameters have been tuned to make the final state distributions agree with fixed target DIS data and data from PETRA, PEP, KEK and LEP [91].

7.3 Event Generator Programs for DIS

LEPTO

The Monte Carlo program LEPTO [17] is a widely used event generator for deep inelastic scattering. It implements the full leading order electroweak cross sections for the underlying parton level NC and CC processes, for an arbitrarily polarized lepton beam. For fragmentation of the produced partons into observable hadrons, the LUND string fragmentation model as implemented in JETSET is used. The program does not include radiative corrections. As the Callan-Gross relation $F_2 = 2xF_1$ holds well except at low x, the longitudinal structure function F_L, as well as target mass effects, are only included as an additional option for electromagnetic interactions. An interface to the standard structure function parametrization library PDFLIB [12] is provided for the evaluation of the parton density functions that are needed for the cross section calculation. Such parametrizations are only valid for $x > 10^{-5}$ (sometimes $x > 10^{-4}$) and $Q^2 > 4\ GeV^2$, although there are already large differences and uncertainties for $x < 0.01$.

In first order QCD, the gluon radiation, $\gamma q \rightarrow qg$, and boson-gluon fusion, $\gamma g \rightarrow q\bar{q}$, processes appear (the exchanged boson taken to be a γ here). The matrix elements (ME) for these processes are known and can be included. Higher than first order effects have been implemented using parton showers (PS). In version 5 of LEPTO, the user has the choice of either using the first order matrix elements for an exact description of the occurrence of boson-gluon fusion or gluon radiation, or using the parton shower approach in order to describe higher order effects. The first choice neglects higher-than-first-order effects altogether, the second choice does not describe the first order processes exactly. From version 6 onwards, LEPTO includes the option of using the first order matrix elements in order to decide whether a lowest order process, a qg process, or a $q\bar{q}$ process has occurred, after which a parton shower can be generated. This gives the best of both worlds: exact calculations of the first order processes, which are treated with the ME method (e.g. hard gluon radiation), plus an additional softer parton shower for higher-than-first-order processes in the leading log approximation.

ARIADNE

ARIADNE [92] is a program for the simulation of QCD cascades, implementing the colour dipole model. It is as such not a complete event generator, but needs an interface to a program like LEPTO for the simulation of the hard subprocess, and to JETSET for the soft fragmentation of produced partons into final state hadrons.

HERACLES

HERACLES [20] is an ep event generator at the parton level, including the complete first order electroweak radiative corrections from the electron and the quark line.

These include interference of leptonic and quarkonic radiation, and the complete one-loop virtual corrections. For a description of radiative effects at HERA, see section 2.6.3. The program is accurate all over phase space, except at very small y and very large x (where corrections are large and negative, but where processes not simulated in HERACLES ensure that the cross section does not become negative). HERACLES itself generates events at the parton level, no QCD cascades are implemented and fragmentation is not simulated. The DJANGO program provides an interface to LEPTO and JETSET in order to perform these tasks.

7.4 Event Generator Programs for Background Processes

Backgrounds for DIS at HERA consists of background events not related to the ep bunch crossings, such as beam-gas interactions or cosmics, as well as ep interaction processes that simulate DIS events. Since DIS events include all processes where the beam electron scatters into the main detector (or where a neutrino is produced in the case of charged current events), including higher order processes like boson-gluon fusion, this latter background consists of events at low Q^2 (photoproduction), where the real scattered electron escapes undetected in the beam pipe, but where a scattered electron is *faked* in the main detector.

PYTHIA

PYTHIA [93] is a Monte Carlo program that, although initially intended for hadron-hadron collisions, now implements many processes relevant for lepton-lepton, lepton-hadron and hadron-hadron collisions. This implies that PYTHIA is able to simulate DIS processes as well as photoproduction processes, but we will use PYTHIA only for the latter.

PYTHIA is able to handle direct photon processes as well as resolved photon interactions, in which a parton in the photon (due to the anomalous structure function of the photon) interacts with a parton in the proton. The inclusion of this process is very important at HERA energies. The lepton vertex is treated with in the WW approximation, with the option of a better treatment of the scattered electron that is only valid at very small Q^2. PYTHIA implements a parton shower in order to model initial and final state QCD radiation. The string model as implemented in JETSET is used for fragmentation.

Chapter 8

DIS Data Selection

8.1 Performance of HERA in 1992

In 1992, HERA has operated with an electron beam energy of 26.7 GeV, instead of the 30 GeV design value. The proton beam energy was 820 GeV, which is the design value. During most of the running period, the electron and proton beams consisted of 10 bunches each, with typical beam currents of 1-2 mA. In almost all runs, one electron and one proton bunch, the so-called *pilot* bunches, had empty proton and electron buckets as counterparts. These bunches were used for studies of the background of electron and proton beam-gas events, since no real *ep* collisions could occur in these bunches. A schematical sketch of this bunch structure is shown in figure 8.1 [94]. The length of the interaction region, which is almost completely

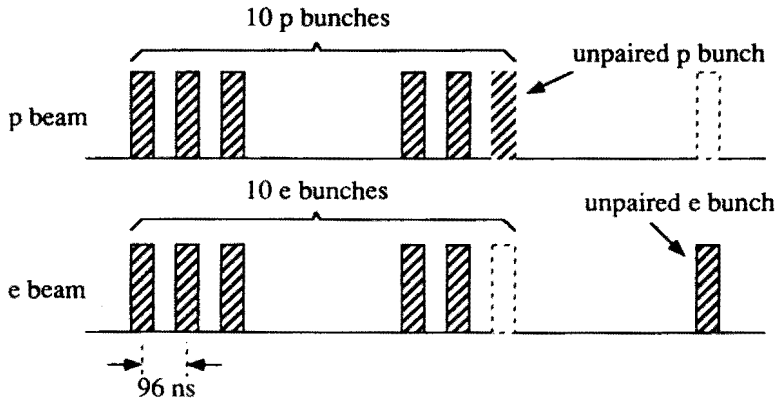

Figure 8.1: *Sketch of the HERA bunch configuration.*

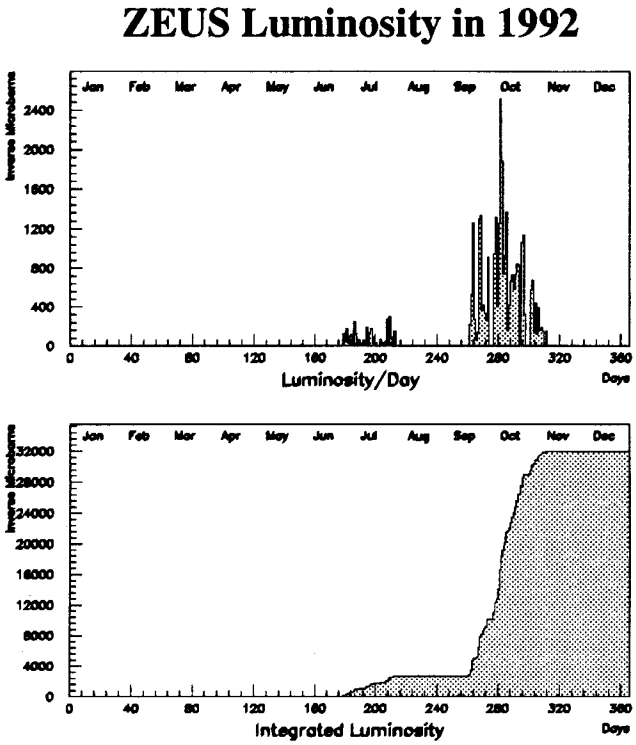

Figure 8.2: *Integrated luminosity in μb^{-1} for ZEUS in 1992, based on on-line data.*

determined by the proton bunch length, was about 40 cm. This is considerably larger than designed, and poses some experimental difficulties in the measurement of the final state for those events in which the actual position of the vertex can not be determined. The proton beam lifetime during luminosity running was typically 50 hours, whereas new electron bunches were injected every few hours.

In 1992, data was taken by ZEUS in 2 periods. Figure 8.2 shows the integrated luminosity for ZEUS in 1992 as a function of time. The estimate shown in this figure is based on on-line luminosity data, and is accurate to approximately 10%. The integrated luminosity for ZEUS in the first running period, from June to the summer shutdown in August, was 2.1 nb^{-1} [1]. After the summer shutdown, careful beam tuning and increasing beam currents increased the luminosity with approximately a factor 10. As a result, ZEUS collected in the second running period from September to early November, data corresponding to an integrated luminosity of 25.8 nb^{-1}, leading to a total integrated luminosity in 1992 of 27.9 nb^{-1}. Note that this is more

[1] These numbers only include data taken with CTD in operation, and only this data has been used in this analysis.

than a factor 1000 below the designed annual integrated luminosity, which implies that the first analyses of HERA physics will be aimed at the region of low x. The statistical error on the luminosity is negligible, the systematic error is estimated to be approximately 5% [95].

8.2 Status and Performance of the ZEUS detector

Not all parts of the ZEUS detector were fully operational in 1992. The CAL, BAC, LUMI, and muon chambers operated as designed. The components FTD, RTD, TRD, and LPS did not take part in the data taking at all. The remaining components were working only partially, or differently from the design.

It was realized in August 1991 that, in the present detector configuration, the designed magnetic field of 1.8 T would induce a force on the superconducting coil which exceeds the design specifications. This force is essentially due to the asymmetric placement of the coil and the calorimeters, which is a logical consequence of the asymmetry in the beam energies. In order to limit the force difference between FCAL and RCAL, it was decided not to run with full current in the main solenoid. This resulted in a central magnetic field of 1.43 T in the 1992 ZEUS operation.

The vertex detector VXD initially did not take part in the data taking, but entered during the second running period. Its data has not been used in this thesis.

The central tracking detector CTD did not have the flash ADC's for the $r-\phi$ readout system installed. That means that all CTD data had to be recorded with the z-by-timing system. This system is installed only on superlayers 1, 3 and 5 of the CTD, and was designed to give an independent z-coordinate measurement with a limited accuracy, the main z measurement coming from the stereo layers. The resolution in z with this system has been shown in the 1992 data to be approximately 4.5 cm. Coordinates in $r-\phi$ can also be measured with this system, but with a limited point resolution of 1 mm instead of the designed 120 μm. It is clear that this has consequences in the pattern recognition and track finding efficiency. The momentum resolution for tracks is degraded with a factor between 10 and 50, and the sign of the tracks can only be determined up to energies of 2 to 7 GeV, depending on the length of the track. In the analysis of the data, the tracks reconstructed in the CTD have been used in order to find an event vertex, but not otherwise.

In the RCAL, only 3 modules were equipped with diodes for the RHES, and these were read out in data taking. The equipped modules were numbers 13 to 15, located next to the beam pipe. The covered region corresponds to $10 < x < 70$ cm in global ZEUS coordinates, with the full vertical EMC length of the modules covered. These RHES diodes operated as specified, and the results have been used in the analysis as an independent check on the reconstructed position of scattered electrons in those modules.

In operation, the components have experienced in general few problems. All channels in the calorimeter were constantly monitored, and the number of bad

channels in the CAL remained limited to a few percent of the total. One operational problem occurring in the calorimeter was the phenomenon of *sparks*. These are thought to be caused by discharges in the PMT base or between PMT cathode and shielding, and cause a non-zero measured energy, at a random time, in a single channel. The majority of these sparks occurred in BEMC channels. They are not difficult to recognize off-line, as a potentially large energy deposit at a random time, with the other side of the cell empty, and a very limited amount of energy in the rest of the calorimeter. However, the rate was high enough to cause a potential problem in the trigger, since the first level trigger was unable to recognize sparks and cut them away. In 1992, algorithms running in the second and third level trigger processors have been sufficiently able to suppress them.

8.3 Trigger Logic

Also the ZEUS trigger system was not complete in the 1992 data taking. Since HERA operated with only a fraction of the designed beam currents, the beam related backgrounds were significantly lower than expected with the full currents. Therefore the available trigger was well able to keep pace with the beams in order to select good physics, and was gradually upgraded as the luminosity increased. A part of the reduction of beam-gas events was obtained using timing information of the calorimeter; this will be explained in section 8.5.

8.3.1 FLT

The first level trigger (FLT) made use of information from the calorimeter, the muon chambers, the luminosity monitor, and the C5 beam monitor counters. Of these, the calorimeter information was most important. The calorimeter first level trigger was not complete as compared to the designed system. A modified system was set up instead. The calorimeter readout was grouped into 448 non-overlapping trigger towers, constructed by summing the signals of 2 adjacent normal calorimeter towers, EMC and HAC separately. The energy sums in these trigger towers were compared on a tower by tower basis to a programmable threshold. The programmable energy threshold was not set constant over the whole calorimeter, but varied between a number of regions. The RCAL was divided into two such regions, one consisting of towers adjacent to the beam pipe, and one with the other towers. FCAL was subdivided into 3 such regions, one adjacent to the beam pipe, one consisting of a ring of towers outside of the beam pipe region, and one with the outermost towers. The thresholds used for the trigger towers in these regions are given in table 8.1. The BEMC threshold was lowered during the second running period in fall 1992 to 1 GeV in order to increase the efficiency for photoproduction events; this has a negligible influence on the efficiency for DIS events.

Whenever the energy measured in any of the trigger towers in EMC or FHAC exceeded the threshold, the event was accepted. If a BHAC trigger tower exceeded

CAL FLT thresholds (GeV)			
section		EMC	HAC
FCAL	beam pipe	50	70
	inner	20	25
	outer	10	10
RCAL	beam pipe	10	x
	outer	2.5	x
BCAL		2.5	1

Table 8.1: *CAL FLT trigger tower energy thresholds in GeV for the different calorimeter regions.*

the threshold, a coincidence with a trigger from the barrel muon chambers was required. The RHAC trigger towers served as veto: if in an event the threshold value of 30 GeV in RHAC trigger towers had been exceeded, the event was vetoed, since this is kinematically not possible for an actual *ep* interaction. First level triggers were vetoed if their signals were in coincidence with a signal from the C5 counters that was in time with the proton beam. This reduces the background from beam-gas interactions, with a negligible loss of DIS events.

Additional implemented triggers required a coincidence of energy in the luminosity monitor and energy (with lower thresholds) in the uranium calorimeter; these triggers select photoproduction events and have not been used in this thesis.

The average first level trigger rate was 8 Hz. The selected data at this stage is still dominated by beam-gas interactions, sparks, and cosmics.

8.3.2 SLT

During the 1992 data taking, the component Second Level Trigger (SLT) systems of CAL and CTD have been tested and commissioned, but the SLT reconstructed clusters have not been used in the trigger. In the summer data taking, the SLT was not used at all. In the autumn, the SLT has been used to reject cosmics that did not give a good track in the CTD, events with a wrong CAL time, and events that were triggered by the low BEMC threshold only, if only very little energy was deposited in the rest of the calorimeter.

8.3.3 TLT

The Third Level Trigger (TLT) ran without major problems on a farm of 6 SGI workstations. In the summer data taking, calorimeter timing algorithms were already available on the third trigger level. For the summer data, a cut on the calorimeter time removed 35% of the events. A spark rejection algorithm removed events that were likely to be caused by sparks. In addition, the TLT had the possibility

to run a track and vertex finding algorithm and a clustering algorithm. These were not used to reject events, but to flag events as probably interesting or background. The final rate of accepted events was approximately 3 Hz per mA of circulating protons.

8.3.4 Efficiency

The trigger acceptance for neutral current DIS events was studied in Monte Carlo simulated events, with full detector and trigger simulation turned on. The trigger acceptance was found to increase with increasing Q^2, and was flat in x. For $x > 3 \cdot 10^{-4}$ and $Q^2 > 8$ GeV^2 it exceeds 97.5%.

The efficiency of the trigger hardware was monitored with selected data samples and with charge injection into the front-end electronics of the calorimeter. Less than 3% of the trigger towers gave no signal, none of which were located in RCAL, which contains the scattered electron in more than 95% of the events of $10 < Q^2 < 100$ GeV^2.

The sensitivity to defects of individual trigger channels was reduced by the fact that a single event can exceed more than one of the trigger thresholds so that the overall efficiency of the hardware was better than 99% in all regions of x and Q^2. The thresholds are low compared to the typical energies of the scattered electrons, thus the trigger calibration scale error (10%) contributed an uncertainty of only ± 1 % to the trigger acceptance for $Q^2 > 10$ GeV^2.

8.4 Recorded Data

In total, $5 \cdot 10^6$ events were accepted by the trigger system and written to tape in the 1992 data taking period. Not all these events correspond to actual *ep* collisions. Moreover, most real *ep* collisions are of the photoproduction type, where the scattered electron disappears down the rear beam pipe. In this thesis, only deep inelastic neutral current scattering events with an electron in the main detector are used. In the remainder of this chapter, the selection of this event sample will be described. In the first step, a data summary tape (DST) is created. This DST contains selected physics events of all kinds, with rather loose cuts on the background. In addition, a sample of triggered events, independent of their origin or whether they have been selected or not, is included for background studies. Events not on the DST are stored on raw data tape, but are not used in physics analysis. From the DST sample, additional and tighter cuts lead to a preselected DIS sample. Finally, some quality cuts are applied to give a final selected data sample.

Before describing the event selection procedure, we will first describe the use of the timing information of the calorimeter in the reduction of upstream beam-gas background.

8.5 Beam-gas suppression by calorimeter timing

Interactions between the proton beam and the rest gas in the beam pipe can occur anywhere in the beam line. The events that deposit energy in the ZEUS detector in general have vertices ranging from inside the main detector ($z = 2$ m) to many meters upstream in the proton beam ($z = \mathcal{O}(-100)$ m). It is clear that particles created in such upstream events can deposit energy in the rear calorimeter, where they enter from the wrong side as compared to good *ep* events. Apart from the possibility of an energy deposit exceeding what is kinematically allowed for *ep* events, such events have a wrong timing. That is, the energy is deposited before the actual bunch crossing appears, whereas the energy in *ep* events must be deposited after the bunch crossing (and therefore the interaction) has occurred. This time of energy deposit can be measured with the calorimeter. As explained in section 4.4.2, the time of the individual pulses in each PMT can be determined. The global FCAL, BCAL or RCAL times t_F, t_B and t_R are then given by a suitable average over the cells in FCAL, BCAL or RCAL, respectively.

The time measured with the calorimeter has a tunable offset which is set such that *ep* events have $t_F = 0$ ns and $t_R = 0$ ns. Beam gas events with energy in FCAL also have $t_F = 0$ ns, but when they deposit energy in RCAL, t_R differs from 0 by the time it takes particles to travel twice the distance RCAL–nominal interaction vertex, and this equals approximately 11 ns. For beam gas events with energy in RCAL, we expect therefore $t_R = -11$ ns, where the minus sign denotes that the signal is too early.

The considerations above hold for the measured average arrival time of scattered particles in FCAL and RCAL. However, in order to actually use the timing in beam-gas suppression, the non-negligible sizes of the proton and electron bunches have to be taken into account as well. For upstream beam-gas events, both t_F and t_R depend on the size of the proton bunch, but the difference $t_F - t_R$ does not. For *ep* physics events, t_F depends on the proton bunch length but not on the length of the electron bunch, and vice versa t_R depends on the electron bunch length but not on the proton bunch length [2]. Since the length of the electron bunch is much smaller than the length of the proton bunch, $t_F - t_R$ depends only on the proton bunch length. The timing can further be affected by a displacement of the bunches, this can be cured by introducing an additional offset that may change from run to run.

The experimental determination of the time in the individual channels requires a minimum amount of energy in the cell in order to obtain sufficient accuracy. The timing resolution is a function of energy, and equals approximately 5 ns at 100 MeV, and 3 ns at 200 MeV, whereas above a few GeV, the resolution is flat at 0.7 ns (this measured resolution includes a contribution from the time jitter of the laser that is used to calibrate the times). The time for a whole calorimeter is determined from the time reconstructed in the individual channels by averaging over many channels, leaving out the channels with insufficient energy. Several methods are in use, on the

[2] In fact t_F and t_R both depend on both bunch lengths, but the effects of the proton bunch length on t_R and the electron bunch length on t_F are negligibly small, as demonstrated in [46].

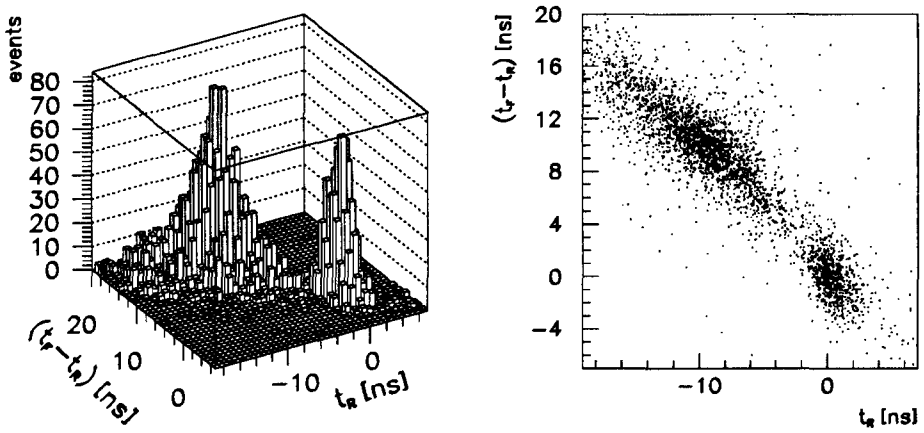

Figure 8.3: *Distribution of the difference between FCAL and RCAL time ($t_F - t_R$) versus the RCAL time (t_R) for DST selected events. The scatter plot at the right hand is the same plot as the lego plot left. For a discussion, see text.*

trigger level a simple average over all channels in the first towers around the beam pipe with an energy exceeding 1 GeV is used. Off-line, an energy weighted average over more channels can be made.

There are two handles on the beam-gas rejection: the absolute RCAL time t_R and the FCAL-RCAL time difference $t_F - t_R$. Figure 8.3 shows the distribution of these two quantities in events taken from an actual *ep* run. There are clearly two separate peaks, the large one from beam-gas events, the small one centered around 0 from good *ep* events.

In addition, the calorimeter timing can be used in the reduction of cosmics, by demanding a BCAL time in agreement with *ep* interactions, or a maximum time difference between upper and lower half of the BCAL.

8.6 Creation of a Data Summary Tape

Deep inelastic neutral current scattering events were selected for a DST in two independent modes. The data on DST is a result of a logical OR of these modes. At this stage, the cuts used to reduce the background are still rather loose, since it is aimed to keep all neutral current deep inelastic scattering events.

Selection mode 1

The first DIS selection mode for the DST consists of the following steps:

1. Reject the event if the signals from the C5 beam monitor are out of the correct time window. This suppresses beam-gas events.

2. Reject the event if the calorimeter timing, as explained in the previous section, does not correspond to an *ep* event. For this, the average times for FCAL, BCAL and RCAL were calculated as an energy weighted average over all cells with an energy exceeding 200 MeV, and with a maximum weight of 2 GeV. If no cells meet this requirement, the time is not calculable. If the times are not calculable, the event is accepted by default. If the times are calculable, but do not meet a set of loose cuts that correspond to what is expected from a good *ep* event, the event is rejected. If FCAL and RCAL times are not calculable, but BCAL time is, a loose cosmics rejection cut is applied. Note that the event must explicitly fail to obey the cuts in order to be rejected.

3. Reject the event if it is not triggered by BEMC or REMC triggers. This demand is justified by the fact that a neutral current DIS event contains a scattered electron in the calorimeter, and by the fact that the trigger thresholds are low enough to trigger on these electrons. If the event is triggered by an REMC trigger, the energy in REMC must exceed 2 GeV. If the event is triggered by a BEMC trigger, the energy in BEMC must exceed 5 GeV. If both triggers fire, $E_{REMC} > 2$ GeV OR $E_{BEMC} > 5$ GeV is required.

4. Reject the event if it is a spark candidate in BCAL. A spark candidate is defined as a recognized electron candidate with an energy larger than 2.5 GeV, consisting of only one cell with an identified bad channel (sparks in cells without a bad channel are much more frequent, but are rejected at trigger level), with empty neighbour cells, and the remaining energy in the calorimeter not exceeding 5 GeV.

5. All remaining events are accepted.

Selection mode 2

The second DIS selection mode for the DST consists of the following steps, of which the first two are equivalent to the first two of selection mode 1:

1. Reject the event if the C5 instrumented collimator gives a veto trigger based on the acceptable time window for *ep* events.

2. Reject the event if the calorimeter timing does not correspond with what is expected from an *ep* event. Only FCAL and RCAL times are used, and the time is calculated as an energy weighted average over all PMT's with an energy exceeding 500 MeV.

3. Reject the event if the total energy in the RCAL exceeds 36 GeV. An RCAL energy larger than 36 GeV (taking the calorimeter energy resolution into account) is kinematically not possible for a real *ep* interaction.

4. Reject the event if $P_l^- < -32$ GeV, where $P_l^- = \sum_{i,\theta>90^\circ} E_i \cos\theta_i$. Again, a failure of this cut is kinematically not possible for a real ep interaction. The sum runs over all calorimeter cells at an angle with respect to the proton beam of larger than 90°, i.e. the rear direction.

5. Reject the event if $\delta = E - P_z = E - \sum_i E_i \cos\theta_i = \sum_i E_i(1-\cos\theta_i) < 5$ GeV. This cut will be explained in detail in the next section. For the calculation of this quantity in the DST event selection, a vertex position of $(0,0,0)$ is assumed.

6. Reject the event if no electron with an energy above 4 GeV could be found. An electron must be found by at least one of two different electron finding algorithms, as described in section 6.15.

7. Reject the event if it is a spark candidate. A spark candidate is defined as an event where the number of cells belonging to the electron candidate is only 1, and the total remaining energy in the calorimeter does not exceed 1 GeV.

8. All remaining events are accepted.

Of a total of $5 \cdot 10^6$ triggers, the DST selection accepted 270000 DIS event candidates. Cosmics have been removed in this sample using an algorithm based on the topology of cosmic muon energy deposits in the calorimeter.

8.7 Preselection of neutral current events

On the DST selected events, a number of additional or tighter cuts are applied. The background at this stage is still high, and consists of beam-gas interactions, cosmics, and photoproduction events. We will describe the next cuts in some detail.

In the evaluation of kinematical variables like P_l^- and P_z, the reconstructed vertex position in z is used. This will increase the accuracy of the calculation, and is especially useful since the large proton bunch length implies a long interaction point. The interaction vertex is reconstructed from a fit to reconstructed tracks. The fit is considered good when 2 or more tracks contribute to the fit, with a reduced $\chi^2 < 10$. This requirement is met in 53% of the events in the final selected sample. For the other events, the vertex point is set to $(0,0,0)$.

The *preselection* procedure consists of the following cuts:

1. Once again, the timing of the calorimeter must be consistent with an ep interaction. Tighter cuts are used here. An energy weighted time average in FCAL and RCAL is made, where only PMT's with an energy larger than 200 MeV contribute, and the maximum weight is 2 GeV. The applied cuts are:

$$-6 < t_F < 6 \tag{8.1}$$

$$-6 < t_R < 6 \tag{8.2}$$

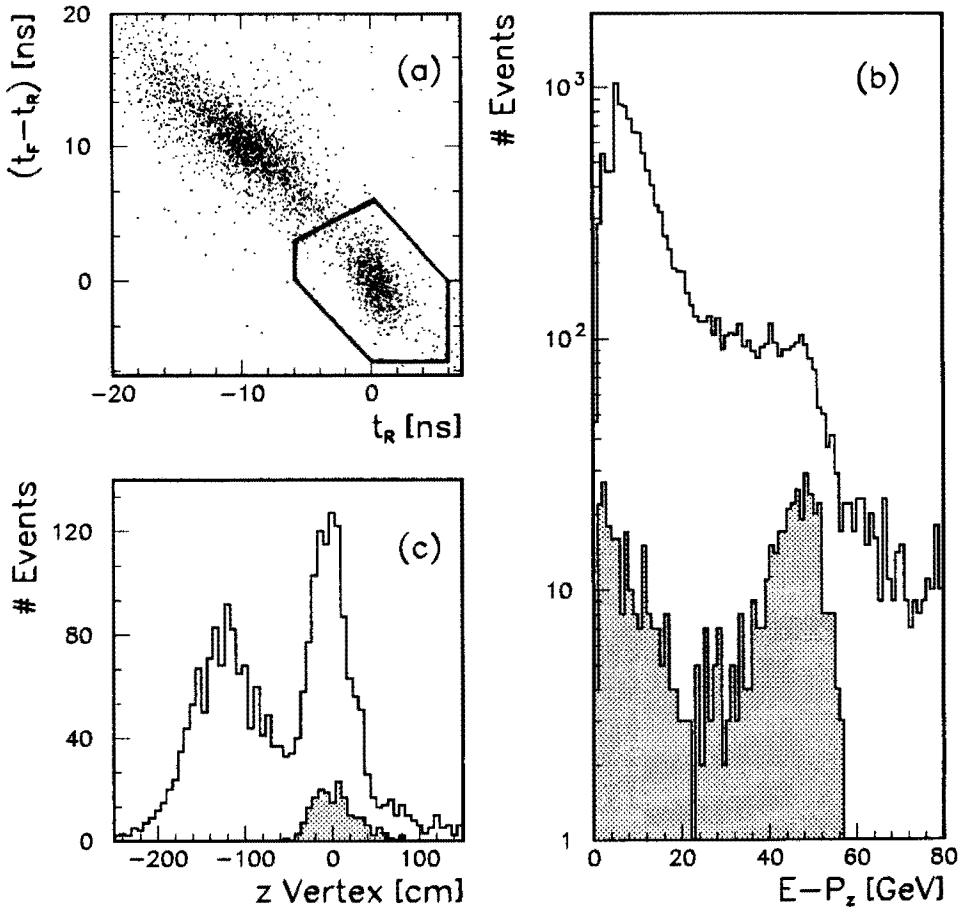

Figure 8.4: *Distributions of various quantities on DST and preselection level, for the summer 1992 data. For a description, see text.*

$$-6 < t_F - t_R < \frac{t_R}{2} + 6 \tag{8.3}$$

with t_F and t_R the FCAL and RCAL time in ns. The latter cut is motivated by the correlation that exists between $(t_F - t_R)$ and t_R. The resulting cut is shown graphically in figure 8.4(a). This figure shows the distribution of $(t_F - t_R)$ versus t_R for DST selected events in the first running period in summer 1992 that had a calculable t_F and t_R time [96].

Events that do not pass the time cut are rejected. When the times are not calculable, the event is accepted by default.

2. Events taken at times when the CTD was not taking part in the data taking

are rejected.

3. Test triggers, such as empty or charge injection triggers, are rejected.

4. Events belonging to empty or pilot bunches are rejected.

5. Spark candidates are rejected. Spark candidates are defined by the most energetic cell having $(E_{left} - E_{right})/(E_{left} + E_{right}) > 0.9$ with neither side in the bad channel list, or an electron candidate with only 1 cell and an imbalance 0. (implying one bad channel).

6. Events with a veto from the veto wall are rejected. It was tested in the data sample that this cut does not cut away good events.

7. Events flagged as beam-gas (type 9) by the TLT are rejected. Also this cut was verified in the data sample.

8. If the LUMI electron calorimeter contains an electron candidate, its energy must not exceed 5 GeV. Since in neutral current DIS events an electron is required in the main detector, no such electron may be found in the LUMI. This cut will eliminate a part of the photoproduction background, namely those events that fake an electron in the detector, but which have their actual electron in the LUMI. Events not passing this requirement are rejected.

9. The quantity $P_z = \sum_i E_i \cos\theta_i$ must be larger than -33 GeV, otherwise the event will be rejected.

10. The energy in the RCAL must not exceed 35 GeV, otherwise the event will be rejected.

11. The quantity $\delta = E - P_z = \sum_i E_i(1 - \cos\theta_i)$ must be larger than 32, but less than 60 GeV. This cut can be understood as follows. We can write $E - P_z$ as:

$$E - P_z = (E - P_z)_{hadrons} + (E - P_z)_{electron} \tag{8.4}$$

with $(E - P_z)_{hadrons}$ equal to $2Ay_{JB}$ according to equation 2.31, and

$$(E - P_z)_{electron} = E(1 - \cos\theta) = 2A(1 - y_{elec}) \tag{8.5}$$

in the notation of equation 2.23. Thus follows:

$$E - P_z = 2A(1 + (y_{JB} - y_{elec})) \tag{8.6}$$

where y_{JB} and y_{elec} denote the variable y from equation 2.4, reconstructed via the Jacquet-Blondel method and the scattered electron, respectively. But since it concerns the same variable y, $E - P_z$ will be approximately identical to $2A = 53.3$ GeV in DIS events, i.e. it is a conserved quantity. $E - P_z$ can exceed 53.3 GeV due to the measurement error (smearing) of the calorimeter,

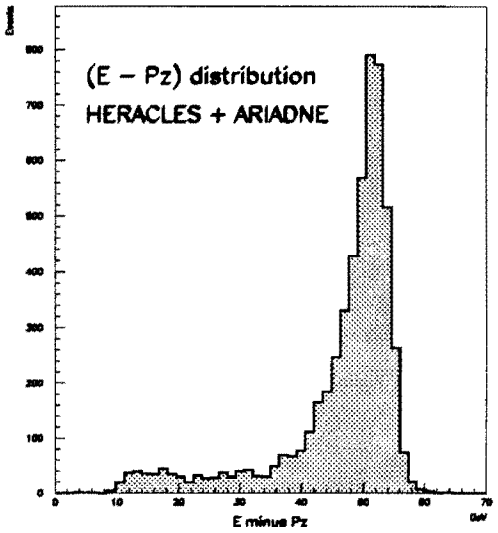

Figure 8.5: *Distribution of $E - P_z$ in Monte Carlo generated DIS events, including radiation.*

whereas it can be considerably lower in events with initial state radiation, where a photon of high energy disappears into the beam pipe and effectively the electron beam energy is lowered. No such problem exists for final state radiation, as long as the photon is measured in the calorimeter. Figure 8.5 shows the $E - P_z$ distribution in Monte Carlo DIS events, including radiation. Note also that the reconstruction of y using the scattered electron leads to large errors at low y, whereas the use of the Jacquet-Blondel method can underestimate y when $E - P_z$ is lost in the beam pipe.

The background has a different $E - P_z$ distribution. The above considerations concerning $E - P_z$ do not hold for beam-gas events, which either originate from upstream but deposit little energy in RCAL, or have an interaction vertex inside the detector (where the timing cut is not effective). They also do not hold for photoproduction events where the scattered electron disappears into the beam pipe. In order for these events to be selected, an electron must be faked, for example by a π^0, in the hadronic final state. The probability for this to happen increases for decreasing particle energy. Therefore, faked electrons from photoproduction events have in general a low energy, and are spread throughout the calorimeter. As such, their value of y, reconstructed via equation 2.23 , peaks at high values. In contrast, y reconstructed via the Jacquet-Blondel method will be too low, since hadronic energy is assigned to the electron. With $y_{JB} < y_{elec}$, $E - P_z$ is distributed around lower values than in actual DIS events. A cut in $E - Pz$ effectively equals a cut in $(y_{JB} - y_{elec})$.

Figure 8.4(b) shows the $E - P_z$ distribution of the data taken in the first running period in summer 1992 [96]. The non-hatched histogram shows the distribution for the DST selected events: the data shows a large peak at low values of $E - P_z$, and a shoulder at the position expected for DIS events. The hatched histogram corresponds to events which were selected with the preselection cuts mentioned above, plus cuts 2, 3 and 4 of the final selection cuts that will be described in the next section [3]. It is clear that the remaining photoproduction background is effectively reduced requiring $E - P_z > \mathcal{O}(30)$ GeV.

The remaining sample after the preselection stage consisted of 14574 events.

8.8 Final Event Selection

Finally, a number of quality cuts are applied to the preselected data. These cuts ensure a good measurement of the final state (in particular the electron), and limit the remaining beam-gas and photoproduction background. These cuts are:

1. $\delta = E - P_z > 37$ GeV, and $y_{elec} < 0.7$ where y_{elec} represents y reconstructed using the scattered electron via equation 2.23. The cut on $E - P_z$ implies a small tightening of the preselection cut on $E - P_z$. The cut on y_{elec} can also be understood from the considerations presented in the previous section: fake electrons from photoproduction events in general have a high y_{elec}. These cuts have been chosen after a comparison of the full preselected sample with Monte Carlo data. Of the preselected sample, 59.7% of the events survive this cut.

2. The energy in FCAL must be larger than 1 GeV. This cut aims to remove remaining background from electron-gas events, where an electron scatters on remaining gas in the beam pipe and is measured in the detector. In such events, FCAL is generally empty. Of the preselected sample, 85.1% of the events survive this cut.

3. Electron finder 2, as described in section 6.15, must find an electron with an energy of at least 5 GeV. Candidates found in FCAL are discarded, as no events are expected there with current luminosity, whereas the background is very high. The position is determined from the calorimeter only, with methods as described in section 6.12. Events with no electron are rejected. Of the preselected sample, 74.5% of the events have an electron with an energy above 5 GeV. It was verified that the reconstructed electron position agrees with the position reconstructed in the HES, when the electron enters the pilot HES modules.

[3] In fact a fiducial cut on the reconstructed electron position around the RCAL beam pipe was used instead of the cut on Q^2; these are, however, almost equivalent.

4. The reconstructed value of Q^2, using the 'double angle formula' of section 2.7, must exceed 10 GeV^2. This ensures that the final state electron can be well measured. This is the case in 58.6% of the preselected events.

The majority of the preselected events that do not survive in the final data sample fail the cuts on $E - P_z$ and y_{elec}, and the cut on Q^2. Many events that do not have the correct $E - P_z$ also have no electron, as expected from photoproduction background events. The cuts on $E - P_z$ and y_{elec} remove also a small part of good DIS events, based on Monte Carlo we estimate this to be approximately 5%.

8.8.1 Resulting Sample

The final sample after application of all cuts consists of 3449 events.

Figure 8.4(c) shows the distribution of the z position of the reconstructed vertices of the accepted events of summer 1992. Again, the non-hatched histogram corresponds to the DST selected events, and the hatched histogram to the events that passed the preselection cuts and the final selection criteria [96]. Only events with a vertex close to the position of the interaction point were selected. The vertex distribution of the full selected sample has a similar shape as shown for the summer data, with a width of 23 cm, in excellent agreement with the width expected from the proton bunch size and with Monte Carlo simulated events. The vertex could be reconstructed, using the criteria mentioned above, in 52.9% of the accepted events, whereas this was the case in 54.5 $\pm$ 1.5 % of the Monte Carlo events (HERACLES+ARIADNE).

For the final sample, the distribution of t_F has a width of 1.6 ns, whereas the distribution of t_R has a width of 0.7 ns. These widths correspond to the expectations based on the time resolution of the calorimeter and the size of the proton bunches.

A subset of the final event sample was scanned visually using the event display program LAZE [52]. No obvious cosmics or beam-gas interactions were found, from which we conclude that the contribution to the sample from these background sources is very small (< 0.5 %). The contribution from cosmics in particular is expected to be very small in view of the cut on the FCAL energy and the requirement that an electron must be found in BCAL or RCAL. The photoproduction background has been studied for the summer data [96] by generating photoproduction events using PYTHIA [93], and passing them through the complete analysis chain. In addition, it is expected that approximately one quarter of these photoproduction events will have an electron with energy greater than 5 GeV in the luminosity monitor. The Monte Carlo studies and the actual search for such electrons agree, and lead to an estimate of the photoproduction background in the summer data of 1.3 $\pm$ 0.7 %. In comparison to the summer data, the trigger efficiency for photoproduction events has been increased in the fall data. The selection criteria used in [96], however, only require $E - P_z > 35$ GeV, and do not use the cut on y_{elec}. Therefore, we do not expect the photoproduction background to the total sample to exceed 2%.

Using a different electron finder (number 1 from section 6.15), 20 events less are selected. All events selected with this second electron finder are also selected with the original one, and the small difference in the number of selected events ensures that the influence of the electron finder on the results obtained with the final sample is very small.

Figures 8.6 and 8.7 show two event display pictures of selected DIS events. Figure 8.6 shows a typical low Q^2 event at HERA, with a clearly identifiable electron, but with a low energy current jet. The electron is scattered over a small angle, and enters in the RCAL, close to the beampipe, in a module equipped with HES diodes. In the bottom right view of figure 8.6, the response of the calorimeter and the HES is shown: the black squares at the left sides of the cells denote an energy deposit in the EMC; the circle shows that the HES has recorded a signal compatible with that of an electron. Figure 8.7 shows a less typical event: the electron scatters into the BCAL, the Q^2 of this event equals 2975 GeV^2 and $x = 0.1$. The current jet is well visible in the FCAL/BCAL overlap region, and balances the electron in p_T.

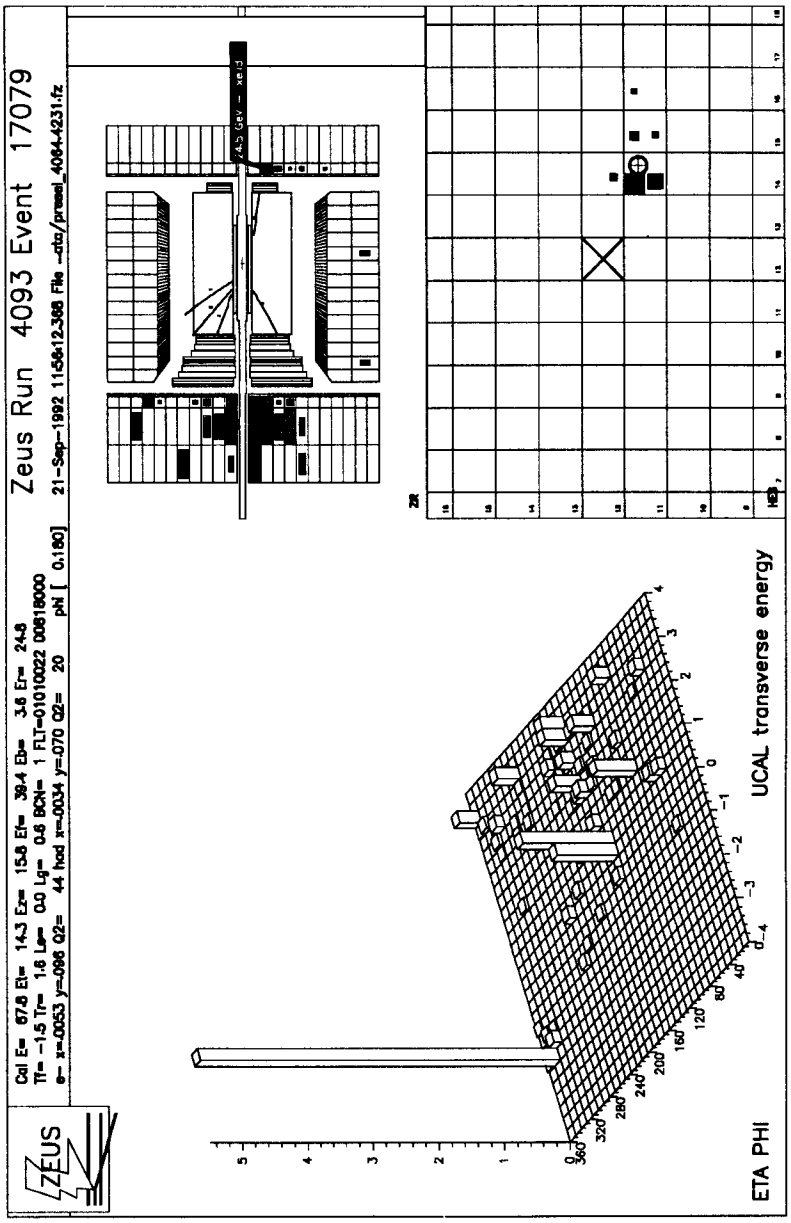

Figure 8.6: *Example of a selected event.*

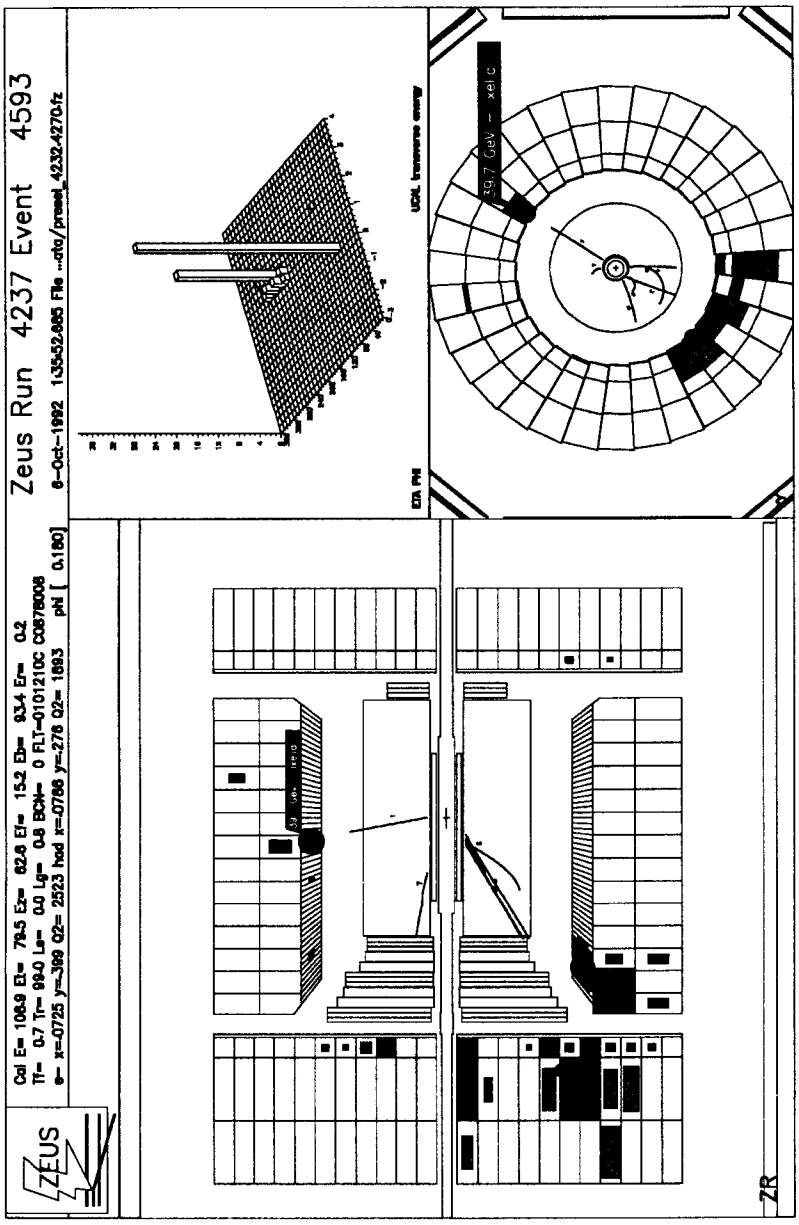

Figure 8.7: *Example of a selected event.*

Chapter 9

Distributions of the Hadronic Final State

In this chapter we will show several distributions concerning the hadronic final state of the selected neutral current DIS events. These distributions include several quantities regarding the hadronic energy flow, and they will be compared to the predictions of several Monte Carlo models. The distributions have been made using the calorimeter only, of other detector information only the reconstructed vertex, if available, has been used. All plots are on calorimeter cell level, no corrections for particles not reaching the calorimeter, or for effects of dead material in front of the calorimeter, have been made. That means that all energies shown here are measured energies, or simulated by the detector simulation program. Information from the backing calorimeter has not been used, since its data and the combination with calorimeter data needs further study.

All data has been reconstructed following the procedures that have been described in chapter 6. For the calorimeter cells, we have used their positions as stored in table CUCELL, described in figure 6.2. The recorded energy in the calorimeter was corrected with UNO data. In order to suppress the noise, energy cuts of 60 MeV have been applied to EMC cells, and cuts of 110 MeV to HAC cells. From the cells, condensates and jets were reconstructed; these objects were subsequently classified with the methods described in section 6.11. The position of reconstructed electrons was determined as shown in section 6.12. For the hadronic energy distributions, the cells assigned to the scattered electron, which was found by electron finder 2 of section 6.15, were removed from the cell list. Using an other electron finder (number 1 from section 6.15) does not significantly change the distributions.

The 1992 data sample of selected events consists of 3449 events, whose selection procedure has been described in chapter 8. Monte Carlo events from several generators were tracked through the detector simulation program MOZART, and selected by the same selection criteria. The various Monte Carlo generators have

been described in chapter 7. We have used for this analysis:

- LEPTO 6, with the option of parton showers matched to the first order matrix elements (implementing the radiation of a hard gluon as well as boson-gluon fusion). This will be denoted as LEPTO ME+PS.
- LEPTO 5, with only the lowest order matrix element for deep inelastic scattering, and with the parton shower option for the simulation of higher order effects. Either W^2 or Q^2 can be chosen as scale for the maximum virtuality of quarks in the parton shower. This will be denoted as PS(W^2) or PS(Q^2).
- ARIADNE 3.1

We have not used the ME option of LEPTO 5, since it has already adequately been shown that the inclusion of first order processes only are not sufficient to describe the data, see also [97] and references therein. Both LEPTO 5 and ARIADNE 3.1 have been interfaced to HERACLES 3.1 for QED radiative corrections; this was not yet possible for LEPTO 6. The Monte Carlo data was generated using the MRS D0 structure function parametrization [15] [1]. However, where not explicitly mentioned otherwise, the MRS D- structure function parametrization has been used in the plots shown in this chapter, by applying an appropriate event weighting technique [99]. We will discuss the influence of the structure function parametrizations on the results.

9.1 Kinematics

The kinematical variables have been reconstructed with the 'double angle formula', which uses only the angles θ and γ, as described in section 2.7. This method has been proven to have a high accuracy, while being fairly insensitive to the calorimeter energy scale [24]. This is an attractive feature, since any miscalibration of the calorimeter energy has severe consequences in the reconstructed kinematics, if, for example, the energy and angle of the scattered electron were used. The hadronic angle γ was reconstructed via equation 2.40. Figure 9.1 shows the reconstructed electron angle θ and the hadronic angle γ distributions for data and Monte Carlo (LEPTO). The distribution of γ is somewhat sensitive to the structure function, whereas θ is influenced to a much lesser extent. The figure shows a fairly good agreement between data and Monte Carlo.

In figure 9.2 the kinematical quantities x, y, Q^2 and W^2 are shown, reconstructed with γ and θ. The Q^2 distribution is strongly governed by the propagator term $\frac{1}{Q^4}$ in the cross section, whereas x and y are more sensitive to the structure function. The agreement between data and Monte Carlo is good.

[1] We have used in this analysis the MRSD parametrizations [15], which are based on preliminary NMC data, and not the MRSD' parametrizations [98], which are based on the final NMC results. The differences between the two are small.

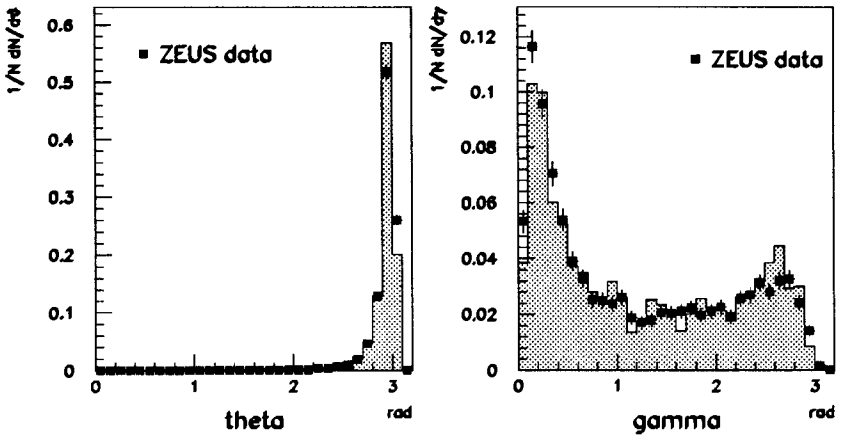

Figure 9.1: *The electron scattering angle θ and the reconstructed quark scattering angle γ in data (points) and Monte Carlo (LEPTO ME+PS MRSD-, histograms).*

The distributions of figures 9.1 and 9.2 for Monte Carlo data are sensitive to the choice of structure function parametrizations. We have tried 4 recent parametrizations: MTB1, MTB2 [14], MRSD0 and MRSD- [15]. Of these parametrizations, we found the distributions made with MRSD- to fit our data best. The parametrizations MTB1 and MRSD0 underestimate the event population at low x (high y, high W^2), MTB1 to a larger extent than MRSD0, whereas MTB2 instead slightly overestimates it.

Although the plots of figures 9.1 and 9.2 have been normalized to an identical number of events, the total cross section of the event sample is also sensitive to the structure function parametrization. Figure 2.4, for example, shows the predicted F_2 for these parametrizations, and F_2 plays a direct role in the total cross section. Our observation of 3449 events at $Q^2 > 10$ GeV^2 in 27.9 nb^{-1} with the described selection criteria implies a neutral current deep inelastic scattering cross section of approximately 130 ± 15 nb at $Q^2 > 10$ GeV^2. The error on this number is dominated by systematics, in particular uncertainties in the luminosity, the efficiency of the selection criteria, and the contribution of the photoproduction background. Without radiative corrections, the various parametrizations predict for this cross section respectively: MTB1: 80 nb, MTB2: 127 nb, MRSD0: 96 nb, and MRSD-: 118 nb. The radiative corrections increase these predictions by approximately 10% [2]. Thus, also the observed number of events is best described by MRSD-.

[2] This number was estimated from HERACLES [20] and [18, 19]. It is relatively small due to the fact that the cut $Q^2 > 10$ GeV^2 removes the events at very low x, where the corrections are largest, and due to the fact that the cut $E - P_z > 37$ GeV cuts out many events with initial state radiation.

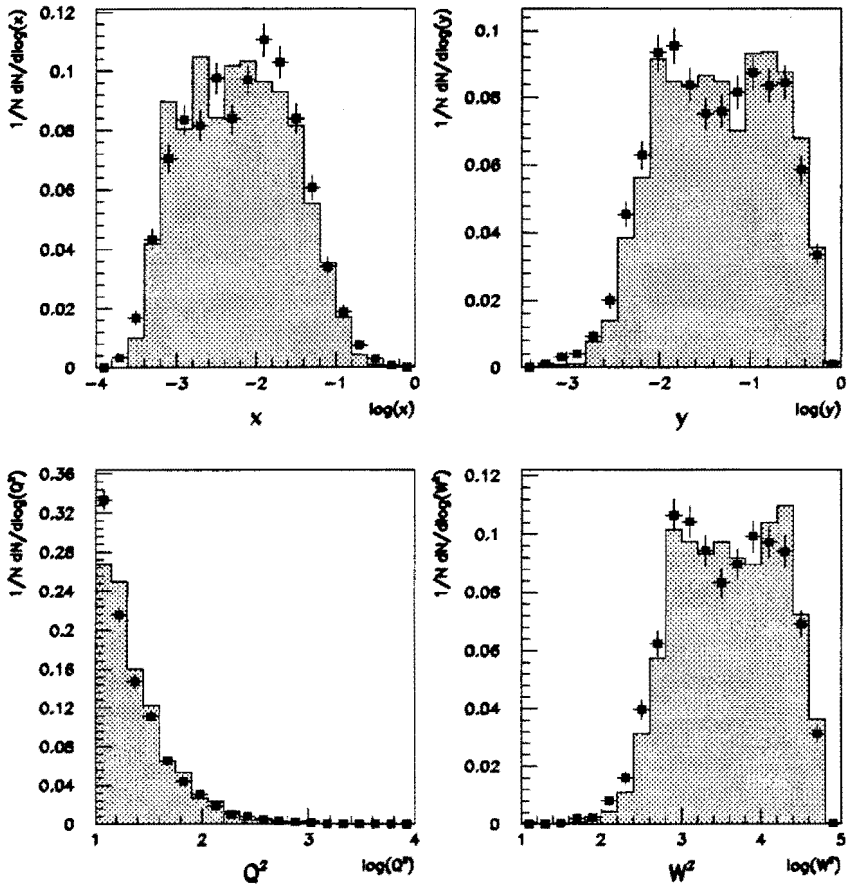

Figure 9.2: *Reconstructed kinematics for data (points) and Monte Carlo (LEPTO, ME+PS, MRSD-, histograms). The double angle formula has been used, all plots are normalized to the same number of events.*

9.2 The Hadronic Energy Scale

The calorimeter energy scale is set before first measurements, using the known conversion factors between reconstructed charge and particle energy (from test beams), and using the UNO currents. The scale is continuously monitored during data taking at HERA by taking regular calibration runs. The correctness of this scale can be tested experimentally in various ways.

The energy spectrum of the scattered electron is in principle known, and depends only little on the structure function. It should give rise to a characteristic peak around the original electron beam energy of 26.7 GeV. When the scattered electron

energy is reconstructed from the measured calorimeter energy, a good knowledge of the composition and distribution of dead material in front of the calorimeter is required. It has been shown [96] that the agreement between ZEUS data and the detector simulation program MOZART (in the version that has been used here) concerning the reconstructed energy of the scattered electron is marginal. It is in particular poor in a number of cells around the rear beam pipe, which have been flagged as a BAD region, and which are not used in this analysis of the energy scale. The problem appears to be a local one, and does not necessarily affect other regions, nor the hadronic energy scale. The reconstructed electron energy has not been used in this thesis in the reconstruction of the kinematics, nor in the plots that follow figure 9.3. The use of the energy of reconstructed scattered electrons is in any case limited, by statistics, to REMC cells in the vicinity of the beam pipe.

A second method consists of the calorimetric measurement of the decay products of particles like π^0, η and ρ. Using reconstructed condensates, ρ particles have already been reconstructed in photoproduction events measured by ZEUS [100]. The use of π^0's is expected to suffer from a large background of random combinations of low energy electromagnetic condensates. Cosmic and beam halo muons have also been used; all these methods, however, are limited to relatively low energy.

A mismatch between electromagnetic and hadronic scale can be detected using the kinematical constraint that the p_T of the scattered electron and the p_T of the hadronic final state must balance. The p_T of the scattered electron was determined from its reconstructed energy and position, excluding those events in which the electron entered the BAD region. The hadronic p_T was determined via a summation over all calorimeter cells not assigned to the electron. The ratio $p_T^{hadron}/p_T^{electron}$ is shown in figure 9.3. The left plot shows this ratio for all selected events with y_{JB} (of equation 2.31) exceeding 0.01, and for Monte Carlo events with the same cut applied. The quantity y_{JB} is relevant in this context, since it denotes the amount of hadronic energy outside the forward beam pipe region, where the target remnant resides. As can be seen in the figure at the right side, for $y_{JB} < 0.01$ the hadronic energy measurement is no longer meaningful. The Monte Carlo simulation by MOZART reproduces the data very well. The ratio is systematically lower than 1 due to the fact that the hadronic p_T is distributed over many particles of low energy, which are less well measured by the calorimeter and suffer from the dead material in front of it. This material, however, appears to be well described by MOZART.

An other method of checking the hadronic energy scale consists of the comparison between kinematical quantities reconstructed using the hadronic energy flow, such as the Jacquet-Blondel method explained in section 2.7, and the same quantities reconstructed with the double angle method, which is, in first order, insensitive to the calorimeter energy scale. A good variable is y, since the influence of the hadronic energy scale on y is immediately clear from equation 2.31. We compare the reconstructed y using equation 2.31 (y_{JB}) to the results using equation 2.36 (y_{DA}) in figure 9.4. This figure shows two profile plots, which show the average values of y reconstructed with the Jacquet-Blondel method in fixed bins of y recon-

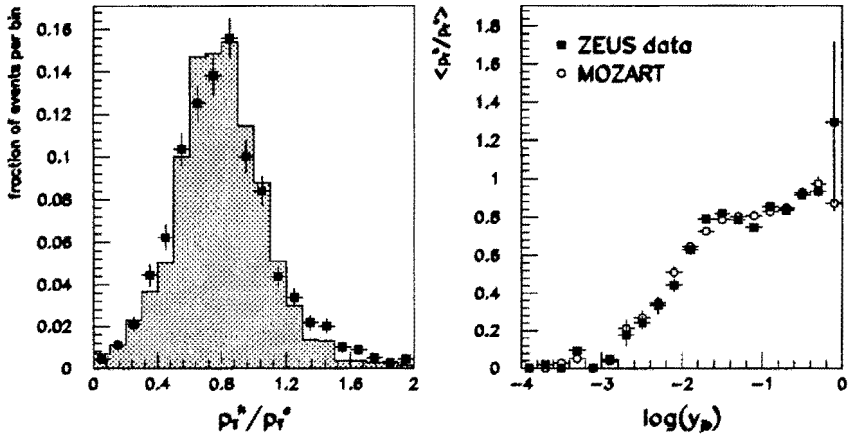

Figure 9.3: *The ratio $p_T^{hadron}/p_T^{electron}$ for data (points) and MOZART (histogram) for all selected events with $y_{JB} > 0.01$ (left), and the average value of this ratio as a function of y_{JB} for data and MOZART (right).*

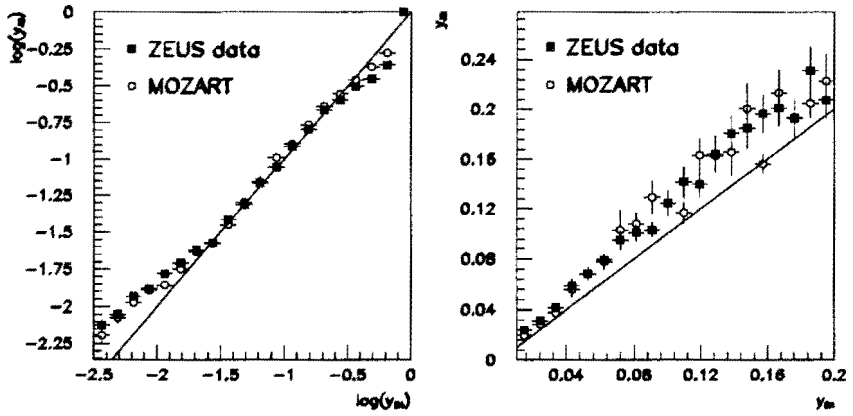

Figure 9.4: *The correlation between y reconstructed with the Jacquet-Blondel method (vertical axis) and y reconstructed with the double angle method (horizontal axis), plotted as two profile plots: with logarithmic scale (left), and with a linear scale between 0.01 and 0.2 (right).*

structed with the double angle method. The black points are ZEUS data, the open points results using MOZART. The left plot of figure 9.4 shows the correlation between y_{JB} and y_{DA} on a logarithmic scale between $10^{-2.5}$ and 1, the line represents a perfect agreement between the two. The Jacquet-Blondel method is expected to deteriorate at low y, where the current jet and target remnant jet overlap, and at very high y, where the current jet looses energy in the rear beam hole. These effects are indeed confirmed by the plot, which shows that the data and MOZART agree very well. The region between $y = 0.01$ and $y = 0.2$ is expanded on a linear scale in the right plot, the central part of this region is expected not to be influenced by either beam holes. A small systematic shift between y_{JB} and y_{DA} is observed, but again, MOZART reproduces the data very well.

We observe that the figures give rise to the conclusion that the hadronic energy scale is correct to within a few %.

9.3 Cells and Condensates

Figure 9.5 shows the distributions of the number of calorimeter cells and condensates per event with enough energy to pass the cuts. These cuts are 60 MeV on an EMC cell, 110 MeV on a HAC cell, 100 MeV on a condensate in the EMC only, and 200 MeV on the other condensates. The points are ZEUS data, the solid histogram represents ARIADNE, and the dashed histogram LEPTO ME+PS. In the distributions of the number of cells per event, all cells around the beam pipe ($\theta < 9°$) have been left out. In the distributions of the number of condensates per event, all condensates are included. The average number of cells and condensates per event increases with increasing W^2. This can be explained by the fact that an increasing W^2 implies, like y_{JB}, an increasing amount of hadronic energy out of the beam pipe regions. Since the condensates are designed to correspond, as much as possible, to showers from single particles entering the calorimeter, the average number of condensates can also be interpreted as event multiplicity, measured with a calorimeter. Both ARIADNE and LEPTO describe the data well. Again, we have varied the structure function parametrization: when MTB1 or MRSD0 are used, the number of events with a high condensate count is slightly underestimated.

9.4 The Hadronic Energy Flow

The distribution of the energy measured in the calorimeter is shown for the selected data sample in figure 9.6. The plot left shows the total energy per event in the calorimeter, when all cells are included. In the right plot, all cells around the forward beam pipe ($\theta < 9°$) have been discarded, in order to obtain a measurement without the target remnant. The points are ZEUS data, the histograms represent the predictions of several Monte Carlo generators. The solid histogram represents ARIADNE, the dashed histogram LEPTO ME+PS, the dotted histogram LEPTO

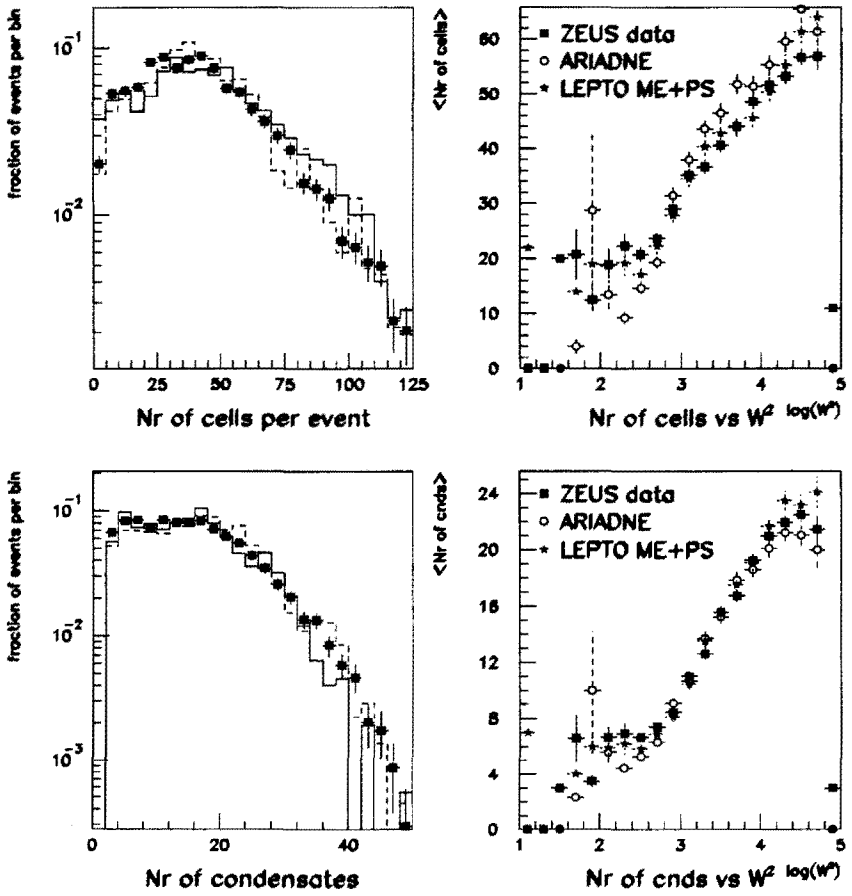

Figure 9.5: *The number of cells and condensates (left) per event. The points are ZEUS data, the solid histogram is ARIADNE, the dashed histogram is LEPTO ME+PS. Also the evolution of this quantity with W^2 is shown (right).*

PS with scale W^2 (PS(W^2)), and the dash-dotted histogram LEPTO PS with scale Q^2 (PS(Q^2)). Changing the structure function parametrization from MRSD- to MRSD0 has only minor influences on the plots. The left plot, which contains all cells, shows that LEPTO ME+PS and ARIADNE both underestimate the total energy. However, since they agree with the data in the right plot, which contains only the cells at $\theta > 9°$, it is clear that this discrepancy is limited to the forward region only, i.e. the target remnant description in these generators slightly deviates from the data [3]. The increase of energy shown by the data pushes the points up

[3] It has been observed [97] that this discrepancy is much smaller in a newer version of ARIADNE, which explicitly contains contributions from boson-gluon fusion.

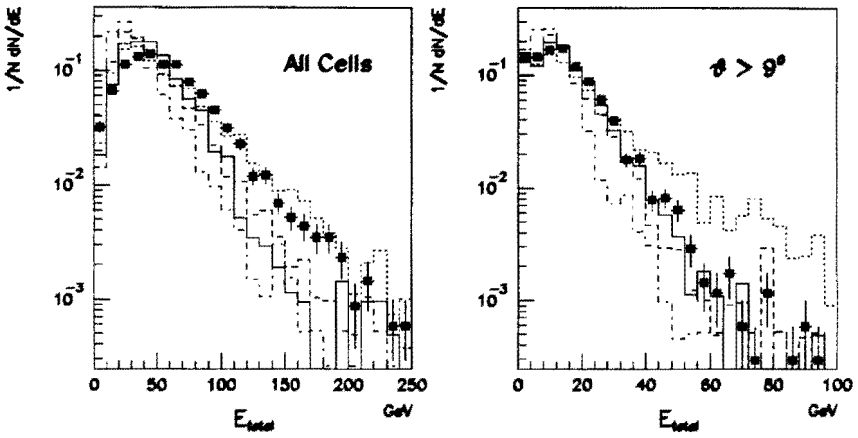

Figure 9.6: *The total energy per event, measured in the calorimeter, including all cells (left) or only the cells at $\theta > 9^\circ$ (right). The points are ZEUS data, for an explanation of the histograms, see the text.*

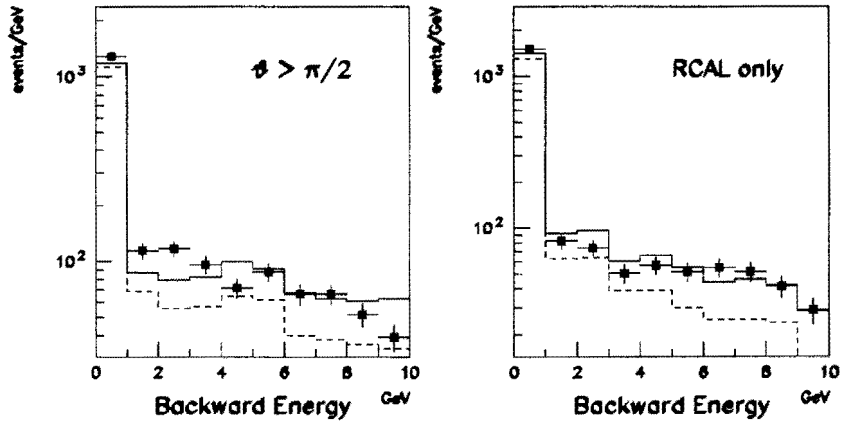

Figure 9.7: *Distribution of energy deposited in the backward region: $\theta > \pi/2$ (left), or in RCAL (right). Points are ZEUS data, solid line: LEPTO ME+PS MRSD-, dashed line: LEPTO ME+PS MRSD0. The plots are normalized to equal luminosity.*

to the line that represents LEPTO PS(W^2). This model, however, fails in the right plot, where the total energy is overestimated. The LEPTO PS(Q^2) model does not represent the data in either plot.

In figure 9.7, the energy deposited in the backward region of the calorimeter is plotted. The left plot shows the energy deposited in the calorimeter at polar angles exceeding $\pi/2$, the right plot shows the energy deposited at $\theta > 2.25$ rad,

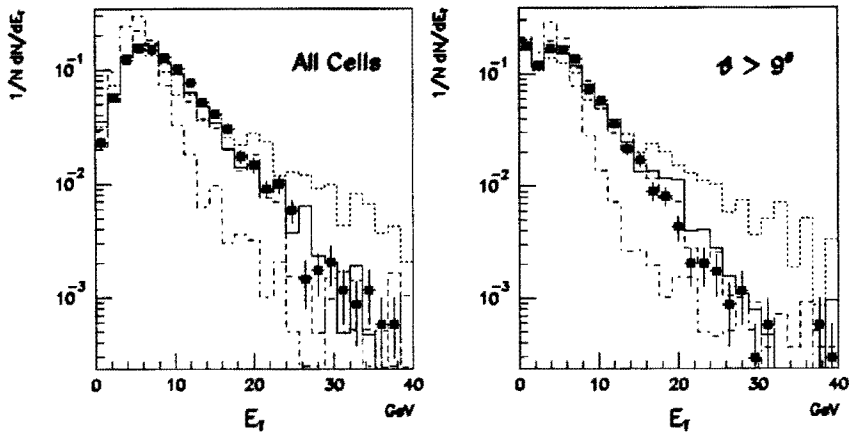

Figure 9.8: *The total E_T per event, measured in the calorimeter, including all cells (left) or only the cells at $\theta > 9°$ (right). The points are ZEUS data, for an explanation of the histograms, see the text.*

which corresponds to the RCAL. The points represent ZEUS data, the solid line the prediction of MRSD-, the dashed line the prediction of MRSD0. The plots are normalized to equal luminosity. These plots are sensitive to the structure function parametrization, and favour MRSD-.

The influence of the target remnant is minimized in the transverse energy. Figure 9.8 shows the scalar transverse hadronic energy, or E_T, distribution. This quantity is defined as:

$$E_T = \sum_{i=1}^{N_h} E_i \sin\theta_i \tag{9.1}$$

where the summation runs over all N_h cells that are hit and have not been assigned to the scattered electron. Again, cells at $\theta < 9°$ have been discarded in the right plot of figure 9.8. The meaning of points and histograms is identical to the previous plot. Both LEPTO ME+PS and ARIADNE describe the data well, whereas both PS models do not: the PS(W^2) model predicts too much E_T, the PS(Q^2) model too little.

An event 'scattering' plane can be defined in the laboratory frame in the following way. In the transverse plane, i.e. the plane perpendicular to the beams, we determine the axis $\hat{a}$ that maximizes the sum $\sum_{i=1}^{N_{cell}}(\hat{a}\cdot\vec{P}_i)^2$, where $\vec{P}_i$ is the projection onto the transverse plane of the cell energy vector $\vec{E}_i$ that points to the cell center and has a length equal to the cell energy. In the evaluation of this quantity, the sum runs over all calorimeter cells, including those assigned to the scattered electron. The event scattering plane is then determined by this axis $\hat{a}$ and the z-axis. In the naive quark-parton model with a single struck quark, this plane corresponds to the scattered electron—struck quark plane. In QCD, gluons are

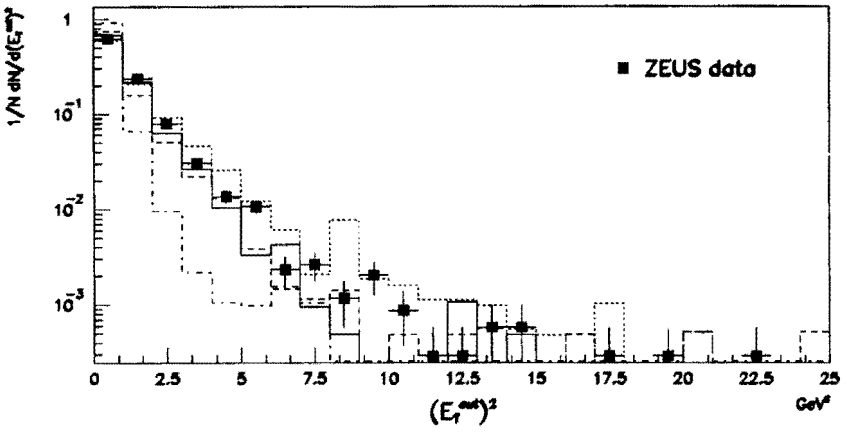

Figure 9.9: *Distribution of the E_T^2 out of the 'scattering plane', per event, measured in the calorimeter. The points are ZEUS data, for an explanation of the histograms, see the text.*

responsible for extra activity in this plane (between struck quark and target remnant), as well as for activity out of this plane, either from radiation from quarks, or from higher-order processes such as boson-gluon fusion. The larger the gluonic activity, the larger will be E_T out of the plane, or $(E_T^{out})^2$.

Figure 9.9 shows the distribution of this quantity $(E_T^{out})^2$ in the selected events. The points are ZEUS data, the meaning of the histograms is identical to those in the previous plots. Again, the data is best described by LEPTO ME+PS and by ARIADNE, and less well by the parton shower models.

It is also interesting to investigate how the hadronic energy is distributed over the calorimeter, as a function of polar angle θ, or pseudorapidity η. The use of pseudorapidity η emphasizes the forward (and backward) region. The hadronic energy flow as a function of θ is shown in figure 9.10. Note that a logarithmic scale is applied on the vertical axis, the actual hadronic energy flow has a deep valley in the BCAL region, and two peaks in FCAL and RCAL, of which the FCAL peak is by far the most dominant. In figure 9.10, cells at $\theta < 10°$ have been excluded in order not to be dominated by the target remnant. The points represent ZEUS data, the solid histogram (with hatched area under the histogram) represents ARIADNE, the dashed histogram LEPTO ME+PS, the dotted histogram LEPTO PS(W^2), and the dash-dotted histogram LEPTO PS(Q^2). In the left plot, the MRSD0 parametrization has been used, in the right plot the MRSD- parametrization. In both cases, the PS models do not describe the data. LEPTO ME+PS and ARIADNE describe the data much better; the MRSD- parametrization tends to overestimate the energy in BCAL (but note the logarithmic scale), but reproduces the peak in RCAL, whereas the MRSD0 parametrization describes the energy in BCAL well, but is not able to reproduce the peak in RCAL. It appears that the data shows a more pronounced

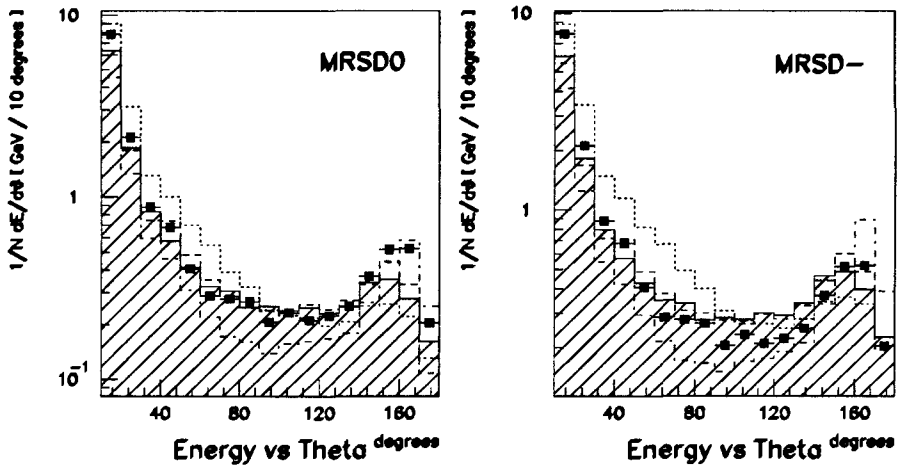

Figure 9.10: *Distribution of hadronic energy versus polar angle θ in ZEUS data (points) and Monte Carlo (histograms), using MRSD0 (left) and MRSD- (right). For an explanation of the histograms, see the text.*

peak than the Monte Carlo.

In these plots, and in many following plots, the event statistics is such that the error bars in the plots, corresponding the statistical errors only, are already smaller than the used marker size.

The hadronic energy flow as a function of pseudorapidity η is shown in figure 9.11 in various bins of x. The points represent ZEUS data, the histograms have the same meaning as in the previous plot. In the left plots, all cells have been included: it is clear that the data shows an excess of hadronic energy in the very forward direction (see the previous footnote). In the right plots, only the cells at $\theta > 10°$ have been included. Moreover, these plots are centered such that the reconstructed value of γ of equation 2.40 corresponds to 0. Since γ corresponds to the struck quark direction in the naive quark-parton model, these plots can be interpreted as showing the energy weighted pseudorapidity gap between cells from the current jet and cells from the target remnant. It is clear that this rapidity gap is largest at low x. The data is well described by LEPTO ME+PS and ARIADNE, again the PS models overestimate (PS(W^2)) or underestimate (PS(Q^2)) the energy.

In the quark-parton model, the event is described as an electron–quark collision, with the rest of the proton as spectator and relatively unaffected. Both struck quark and target remnant fragment and form a jet. Since electron and quark jet must balance p_T, they must be back-to-back in azimuth ϕ. We therefore expect a peak in the azimuthal energy flow in the events, in the opposite direction to the scattered electron. The form of this energy flow peak is influenced by the amount of gluon radiation in and out the jet, and therefore differs between the various models. Figure 9.12 shows the azimuthal hadronic energy flow with respect to the

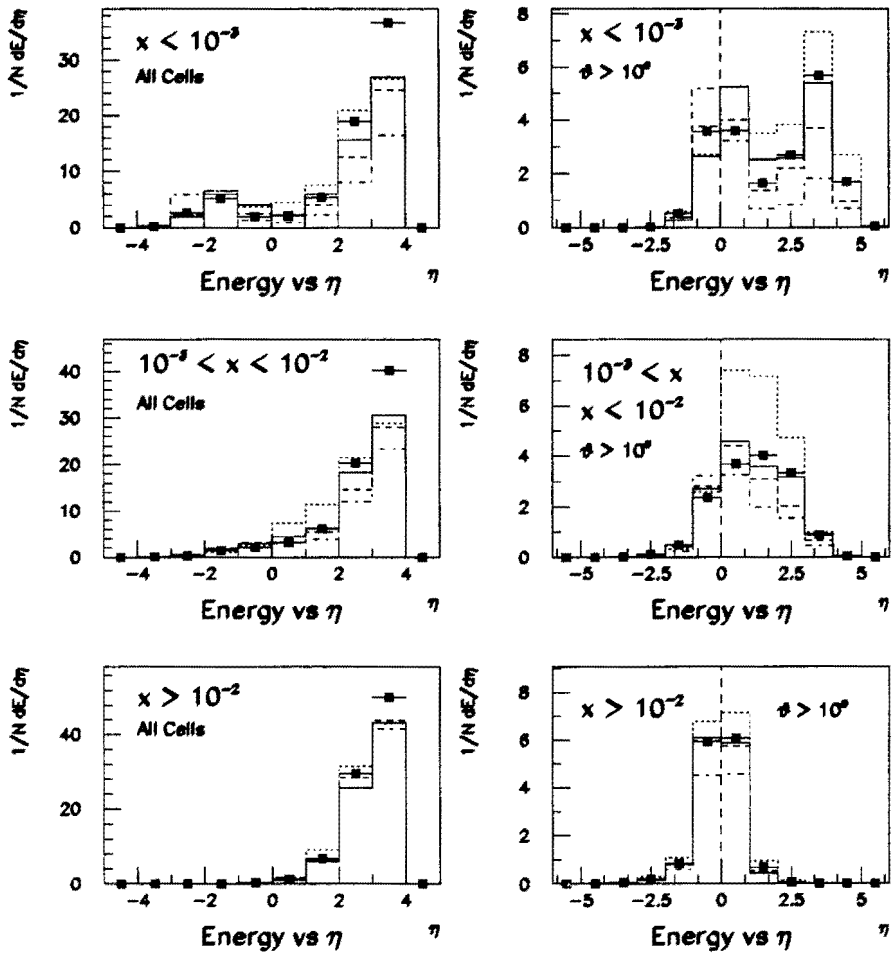

Figure 9.11: *Distribution of energy versus pseudorapidity in ZEUS data (points) and Monte Carlo (histograms), in various bins of x. In the left plots, all cells have been included; in the right plots only the cells at $\theta > 10°$, and these plots are centered around γ.*

direction of the scattered electron, for all events, and in three bins of x. As usual, the ZEUS data is shown as points, and the histograms have the same meaning as in the previous plots. The lowest bin in x suffers from a lack of statistics. Nevertheless, it is clear that the PS models again fail to describe the data, in contrast to LEPTO ME+PS and ARIADNE. The plots also show that the hadronic energy is in general back-to-back to the electron, as expected. The width of the peak decreases with increasing x: at high x the events are more jet-like, whereas at low x, the jet

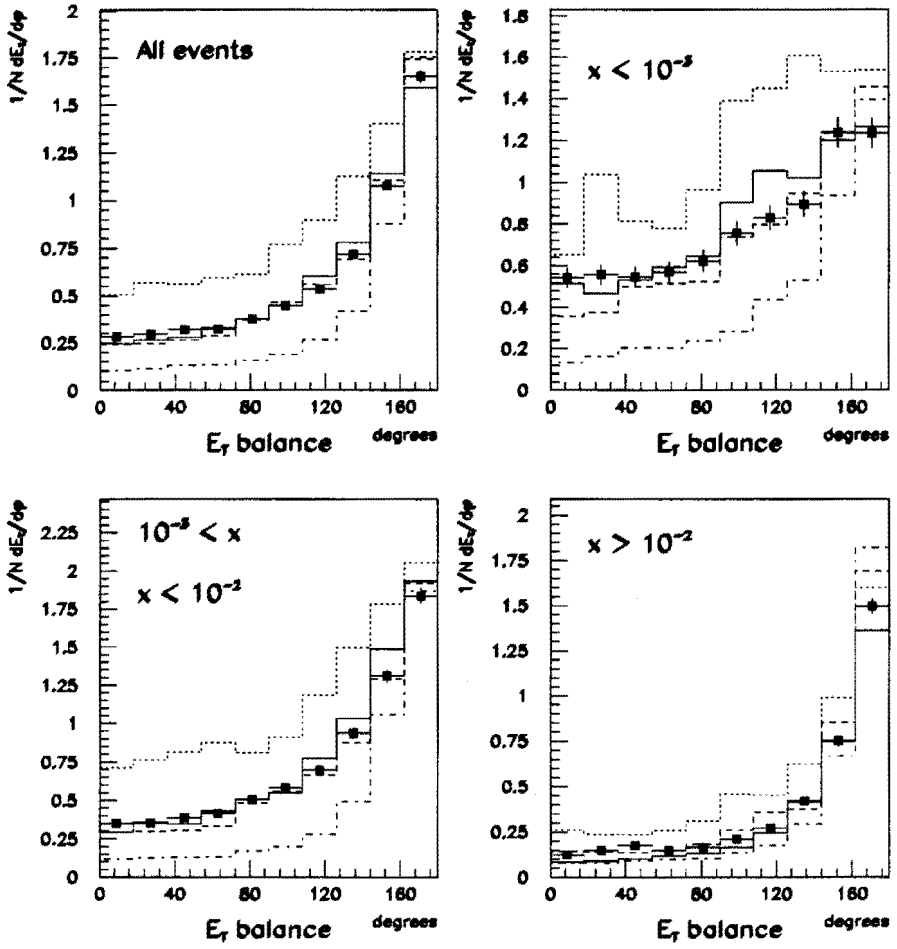

Figure 9.12: *Distribution of E_T in the transverse plane, with respect to the scattered electron, for all events and in three bins in x. The points represent ZEUS data, the histograms various Monte Carlo generators (see text).*

structure is less striking.

9.5 Discussion

From the plots that have been presented in the previous sections we can draw the following conclusions.

The distributions of kinematics, cells hit, and hadronic energy flow, behave largely as expected. Thus, we have reasons to believe that the hadronic energy

scale is correct to the level of a few %. It is shown that the ZEUS calorimeter is able to accurately measure the hadronic energy flow in deep inelastic scattering events at HERA. The behaviour of the detector is in general well modeled by the detector simulation program MOZART. We have also shown that the measurement of the hadronic energy flow can be used for physics analysis.

For those distributions that are sensitive to the choice of structure function parametrization, we find that the MRSD- parametrization describes the data best. This is confirmed by the observed cross section. However, the plots are not fully conclusive, and not all aspects of the data are fully described by one single parametrization. Nevertheless, we conclude that MRSD0 is less effective in its description, and that the data does not agree with MTB1. The MTB2 parametrization does not perform as well as MRSD-, but better than MRSD0 and MTB1. The model behind the MRSD- and MTB2 parametrizations involves a strong rise of the F_2 structure function (and in particular the gluon distribution in the proton) at low x.

The models implemented in LEPTO ME+PS and ARIADNE describe the data best. The parton shower model with scale Q^2 underestimates the energy that has been observed to flow between struck quark and target remnant, and out of the plane of scattering. The parton shower model with scale W^2, in contrast, overestimates this energy. The difference can be understood by the kinematics of the 1992 ZEUS event sample: most events have a low Q^2 and a low x, and therefore a high W^2. Therefore, the predicted amount of gluon radiation, and thus the predicted amount of energy between struck quark and target remnant, and energy out of the scattering plane, is very different. It is shown that neither of these two scales agrees with the data. Matching the parton showers to the first order matrix elements, on the other hand, works very well. Also the colour dipole model is generally well able to predict the hadronic final state at HERA.

The energy observed in the data in the very forward region is not described by LEPTO ME+PS or ARIADNE. This discrepancy has not influenced our previous conclusions, since in most cases we are not interested in this region, and we have left it out. The reason for this discrepancy could lie in a problem in the hadronic energy scale for this region, or could be a problem in the Monte Carlo models. The observation, however, that a more recent version of ARIADNE does describe the data much better [97] leads to the conclusion that the problem does not lie in the hadronic energy scale.

Several systematic checks have been performed within the ZEUS collaboration in this analysis [97]. The predictions of ARIADNE calculated both with and without the interface to HERACLES, which incorporates QED radiative corrections, were compared. No significant differences were observed. Several parameters in the fragmentation and decay package JETSET, notably the energy-momentum fraction and the p_T distribution of the primary produced hadrons, and the Gaussian width of the primordial transverse momentum distribution of the partons within the proton, were varied. Even when these changes were larger than any made to tune the default settings to lower energy data [91], no significant changes were observed. The same holds for variation of the *minimum* virtuality in the initial and final state parton

showers, and the QCD parameter Λ_{QCD}.

Chapter 10

Jets

In this chapter, we will try to look for jet structure in the data. After discussing the use of γ as the direction of the current jet in lowest order processes, we will apply the cone jet finding algorithm to the data, and compare to the Monte Carlo. Apart from the individual jet distributions, we are particularly interested in events with more than 1 jet (not counting the target remnant, nor the scattered electron), since these may be evidence for the occurrence of higher order QCD effects, such as boson-gluon fusion or QCD-Compton scattering.

10.1 Evidence for jet structure

In the previous chapter, the hadronic energy flow as a function of polar angle θ and pseudorapidity η, and the transverse hadronic energy flow with respect to the scattered electron have been presented. These distributions showed the behaviour as expected from the kinematics of deep inelastic scattering: an increase of energy flow back-to-back in azimuth ϕ with respect to the scattered electron, and an increase in the energy flow in θ or η outside of the target remnant region, whose location and magnitude depends primarily on x. This energy flow can be interpreted as the current jet, balancing the electron in p_T.

In section 2.7 we have introduced the variable γ as the polar angle of the outgoing struck quark in the quark-parton model. In equations 2.39 and 2.40 we have indicated how γ can be reconstructed from the measured hadronic final state. In order to show how well γ represents the direction of the current jet, we have plotted the hadronic energy flow around γ in figure 10.1. In this figure, the average hadronic energy per event is plotted for the selected data sample, against pseudorapidity. In these plots, the reconstructed value of γ always corresponds to 0. Four bins in x are chosen: $x < 10^{-3}$, $10^{-3} < x < 10^{-2}$, $10^{-2} < x < 10^{-1}$, and $x > 10^{-1}$. At low x, the current jet has a low energy, but is well separated from the target remnant. As x increases, this gap in rapidity decreases, but the current jet energy increases, as

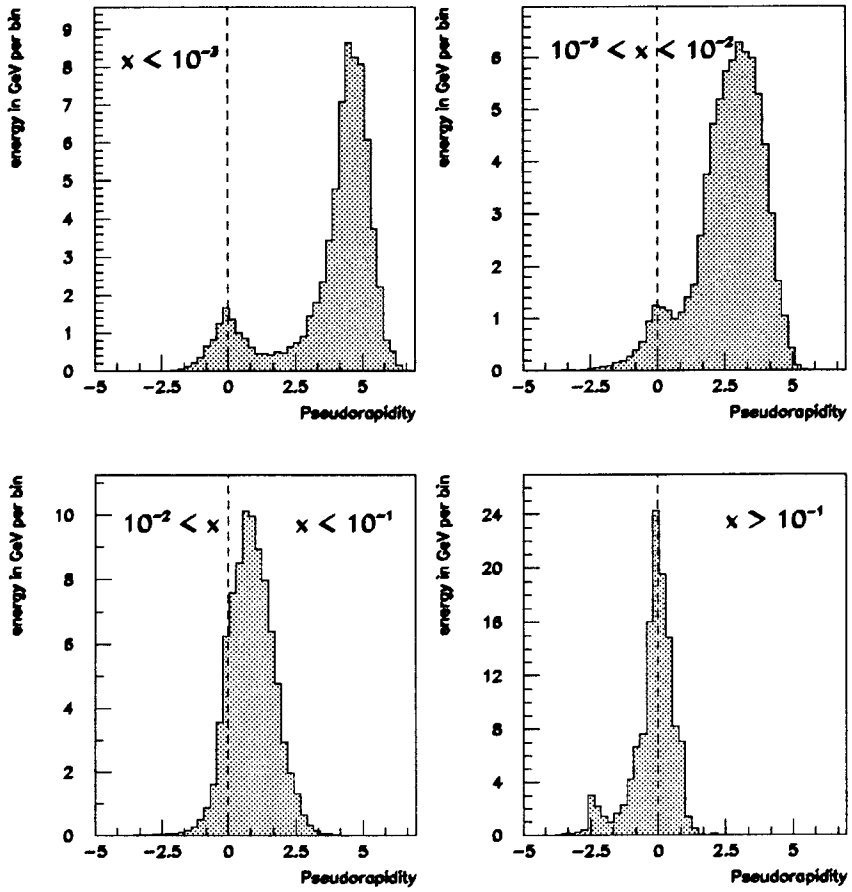

Figure 10.1: *Hadronic energy flow around γ. Plotted is the hadronic energy flow versus pseudorapidity, with reconstructed γ centered at 0, in four bins of x: $x < 10^{-3}$, $10^{-3} < x < 10^{-2}$, $10^{-2} < x < 10^{-1}$, and $x > 10^{-1}$.*

can be seen from the total energy in the histogram bins that correspond to current jet and target remnant. From $x > 10^{-2}$, the current jet begins to overlap with the target remnant, and the latter is only seen as an asymmetry in the peak. It is clear that γ represents the current jet in lowest order processes well. We have tried to reconstruct γ by applying a jet algorithm to the data and explicitly reconstructing a current jet; the resulting accuracy in the reconstruction of the event kinematics is nowhere better, and often slightly worse, than what has been obtained with equation 2.40.

10.2 Jet multiplicities

It has thus been shown that the variable γ as reconstructed in the data well represents the current jet, as it exists in the lowest order graphs corresponding to the naive quark-parton model. However, since γ effectively averages over the whole hadronic final state, it is not useful for reconstructing the individual jets in higher order processes. This applies to higher order processes in deep inelastic scattering, but γ is also useless in photoproduction events. We limit ourselves in this thesis to DIS events, a jet study in photoproduction events has been published in [101]. Multi-jet events thus need to be recognized by applying explicitly a jet finding algorithm to the data.

In this chapter, we have applied the cone jet finding algorithm, as explained in section 6.9, to the selected data sample. We emphasize that this algorithm follows the proposed standard for hadron collider physics as laid down in the 'Snowmass Convention' [76]. We have chosen for the cone algorithm since it was shown, in chapter 6, to follow the parton jets best in energy and angle. As in the study presented in that chapter, the minimum E_T for a cell in order to be regarded as a cluster seed was set to 300 MeV, and the cone radius in $\eta - \phi$ coordinates was set to 1. All cells assigned to the reconstructed scattered electron were removed from the list of cells before the jet algorithm was applied, the set of final state reconstructed jets thus does not contain the scattered electron.

The minimum energy for a reconstructed jet in order to be accepted as a valid jet was set to 4 GeV. Below this energy, the particles that result from the fragmentation of a quark or gluon have such low energy, and may emerge under such different angles, that they do not form a distinct jet. In the selected data sample, 88% of all events were found to have at least one jet of more than 4 GeV, at any angle. This number is well reproduced by LEPTO, and has only very little sensitivity to the structure function parametrization. The observation that not every event contains a jet is due to the fact that a minimum E_T of 300 MeV is required for a cell in order to serve as a seed; in events which have only hadronic activity around the forward beam hole this requirement is often not met.

Due to the fact that the $\eta - \phi$ plane has a singularity for $\theta \rightarrow 0$, the cone algorithm has a problem for the very forward region. Since we are not interested in the target remnant, but only in jets outside of the beam pipe region, we only accept jets at $\eta < 2$, i.e. $\theta > 15°$. This reduces the number of events with at least one jet above 4 GeV to 37%.

The jet multiplicities depend on the minimum energy, or transverse energy, required for a cluster in order to be accepted as a jet. The dependency of the jet multiplicity in data and Monte Carlo on the energy, or E_T cut, is shown in figure 10.2. The figure shows the fraction of events with 0, 1 or 2 jets, as a function of the energy or E_T cut applied, for data (circles) and Monte Carlo (lines). In the figure, LEPTO ME+PS is represented by a dashed line, ARIADNE by a dotted line with small spacing, LEPTO PS(W^2) by a dash-dotted line, and LEPTO PS(Q^2) by a dotted line with large spacing. The data is best described by LEPTO ME+PS,

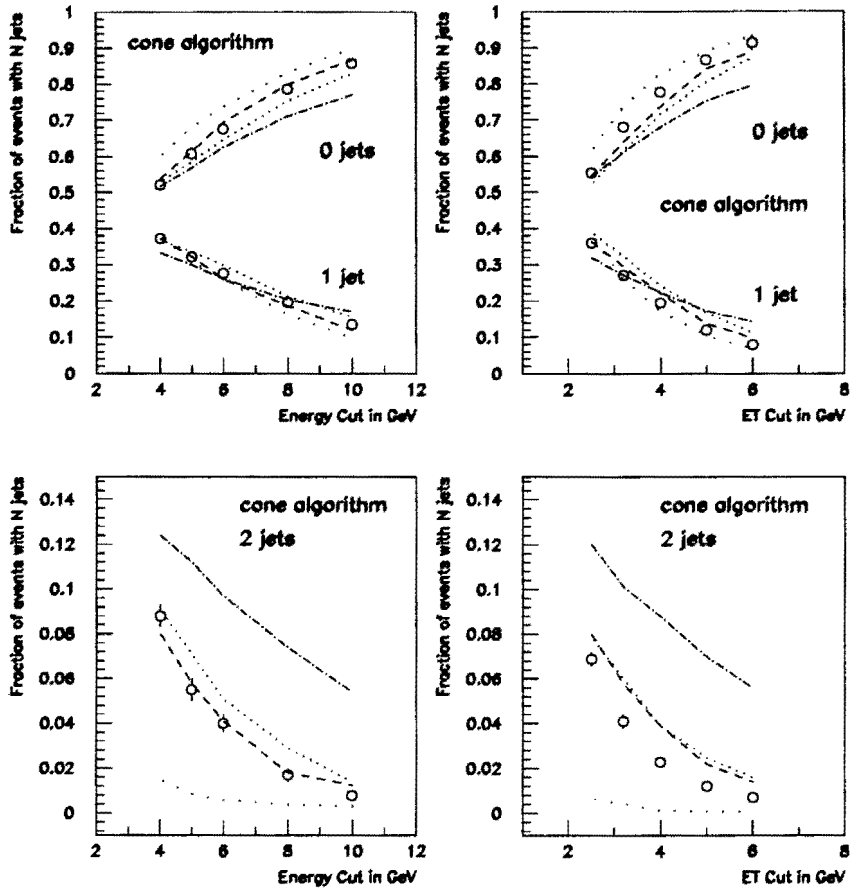

Figure 10.2: *Jet multiplicities per event as a function of the minimum energy (left) or transverse energy (right) required for a jet. The open circles represent ZEUS data, the lines various Monte Carlo models explained in the text.*

and to a lesser extent by ARIADNE. This holds in particular for the jet multiplicity versus the energy cut; when the E_T cut is applied, the amount of jets, and in particular two jets, is slightly overestimated. This effect can thus be attributed to a small difference in reconstructed angle between data and Monte Carlo (see next section), which is a consequence of the energy surplus shown by the data in the forward direction. The parton shower models with scale W^2 or Q^2 do not describe the data.

The jet multiplicities are not significantly affected by the choice of structure function parametrization.

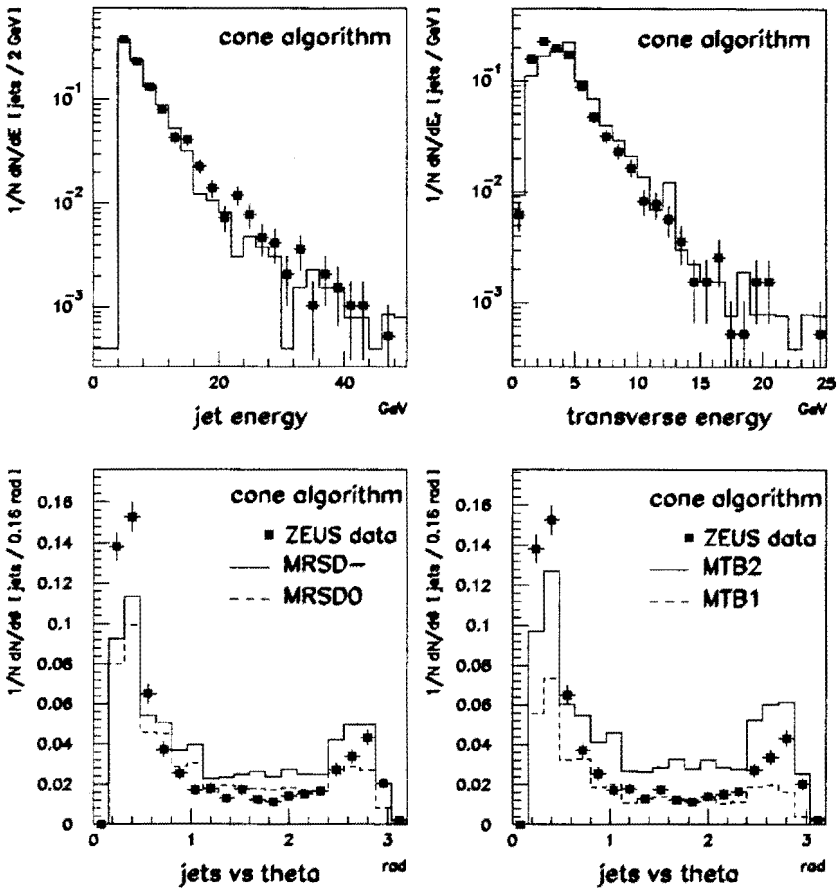

Figure 10.3: *Energy, transverse energy, and reconstructed polar angle θ, for reconstructed jets with energy > 4 GeV, and $\eta < 2$.*

10.3 Jet distributions

Figure 10.3 shows various distributions of the reconstructed jets with energy exceeding 4 GeV, and $\eta < 2$. The top left plot shows the distribution of jet energy, it is well described by LEPTO, as is the transverse energy distribution. In the two bottom figures, the reconstructed polar angle distributions are shown for data (points) and LEPTO with various structure function parametrizations. In the bottom left plot, the solid line represents MRSD-, the dashed line MRSD0. In the bottom right plot, the solid line represents MTB2, the dashed line MTB1. The bottom plots are normalized to correspond to the number of events expected (observed) with equal luminosity. The observation that different parametrizations lead to different numbers of jets for equal luminosity is due to the fact that the number of events itself

is sensitive to the parametrization (see previous chapter), the actual number of jets per event is only marginally dependent on the structure function. We observe that the number of reconstructed jets lies between what was expected with MRSD0 and MRSD-; MTB1 gives not enough jets, MTB2 too many.

The data shows two clear peaks in the jet distribution, namely at very low polar angles and at high polar angles. The MRSD0 parametrization does not describe the sharp rise of the number of jets at high angles; the MRSD- parametrization does much better, but predicts too many jets in the BCAL region. None of the parametrizations completely describe the low angle peak, which is caused by the surplus of energy shown by the data in the very forward region.

The shape of a jet, i.e. the energyflow around the jet axis, has been investigated in [102], and was found to agree with LEPTO ME+PS.

10.4 Multijet production in DIS

As shown in the previous section, we have observed events with more than 1 jet outside the beam pipe region. In this final section we will take a closer look to these events. In order to minimize the possibility of a single jet artificially split in two by the algorithm, we apply the tight criterion of $\eta < 2$ and $E_T > 4$ GeV for the jets. With the described selection and jet criteria, we find that 2.3% of the selected events have two jets. For the data taken in the fall (25.8 nb^{-1}, 3176 selected events), this amounts to 73 events. A few distributions of these events are shown in figure 10.4.

The hatched areas in this figure represent the contributions of the two-jet events in the data sample. The figure shows that the two jet events have a high W^2, and a high E_T, as expected from higher order processes. The azimuthal angle between the two jets peaks at π, so that the two jets are predominantly back-to-back.

The polar angles θ_{min} and θ_{max} of the two jets in the two-jet events are shown in figure 10.5. There is a good correspondence between data and Monte Carlo, but the data again shows an increase of jets at low polar angles, which we believe to be caused by the excess of energy in that region shown by the data, and which is not reproduced by the versions of the Monte Carlo's we have used.

For the fall data, the reconstructed x and Q^2 is shown for all selected events, and for the two-jet events, in figure 10.6.

All two-jet events were scanned with the event program LAZE [52]. No obvious cosmics [1] or beam gas events remained in the sample after scanning.

In [102], LEPTO ME+PS has been used to investigate the origin of the two-jet events. In this Monte Carlo approach, the events can originate from higher order processes like boson-gluon fusion or QCD-Compton scattering, as implemented in the first order matrix elements, as well as from the occasional occurrence of a parton

[1] The original sample of two-jet events contained one event extra, which by scanning was found to consist of a superposition of four cosmic muons, two of them in FCAL and two in BCAL, one of which faked the signal of an electron.

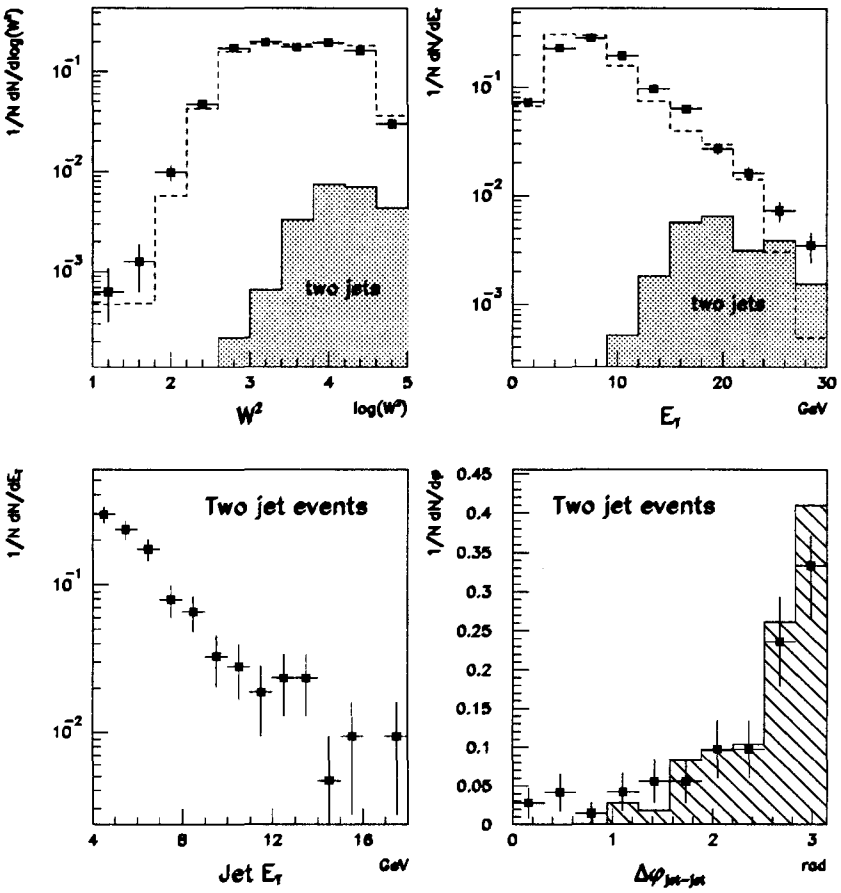

Figure 10.4: *The two jet sample. Top: distributions of W^2 and E_T in data and Monte Carlo, and in the two jet sample (hatched). Bottom: jet E_T distribution, and the azimuthal difference $\Delta\phi$ between the two jets, for data and Monte Carlo.*

with high transverse momentum in a parton shower. Whether such a parton should be counted as a higher order process or product of a parton shower, depends on the particular cut-off scheme used in the Monte Carlo to avoid singularities; in the scheme adopted by LEPTO ME+PS, the contribution of higher order processes like boson-gluon fusion and QCD-Compton scattering dominates above $E_T^{jet} = 8$ GeV.

Figures 10.7 to 10.10 show event display examples of multi-jet events in the deep inelastic scattering sample. In the lego-plots, the scattered electron is characterized by the high peak in few cells; the jets are clearly visible. Figure 10.10 shows a three jet event, which can originate from the radiation of a second hard gluon in a

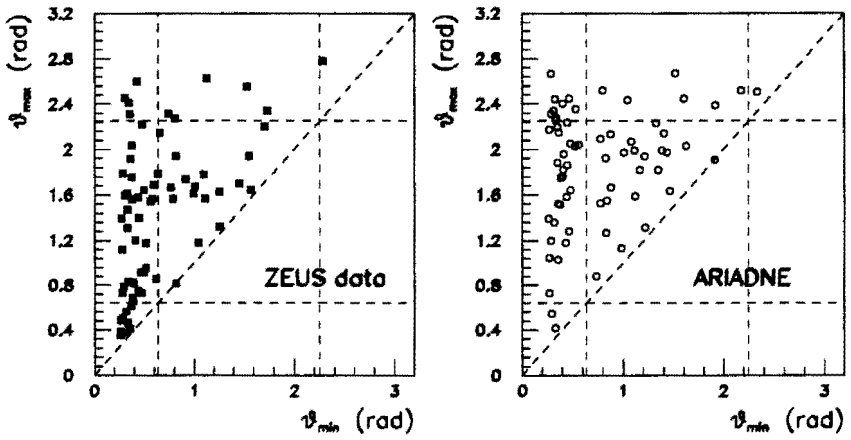

Figure 10.5: *The polar angles θ_{min} (horizontal axis) and θ_{max} (vertical axis) of the two jets in the two-jet events, for ZEUS data (left), and ARIADNE (right).*

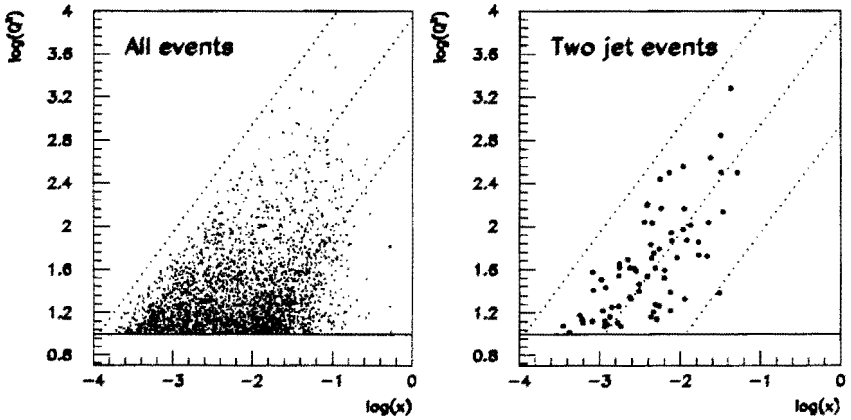

Figure 10.6: *Reconstructed values of x and Q^2 for all selected events (left) and the two-jet events (right).*

QCD-Compton event, or from a gluon radiated from the quark or antiquark in a boson-gluon fusion event.

We conclude that the data indeed shows evidence for multi-jet events in deep inelastic scattering at $Q^2 > 10$ GeV^2, and that a Monte Carlo analysis attributes these events to higher order processes like boson-gluon fusion and QCD-Compton scattering at high transverse jet energies.

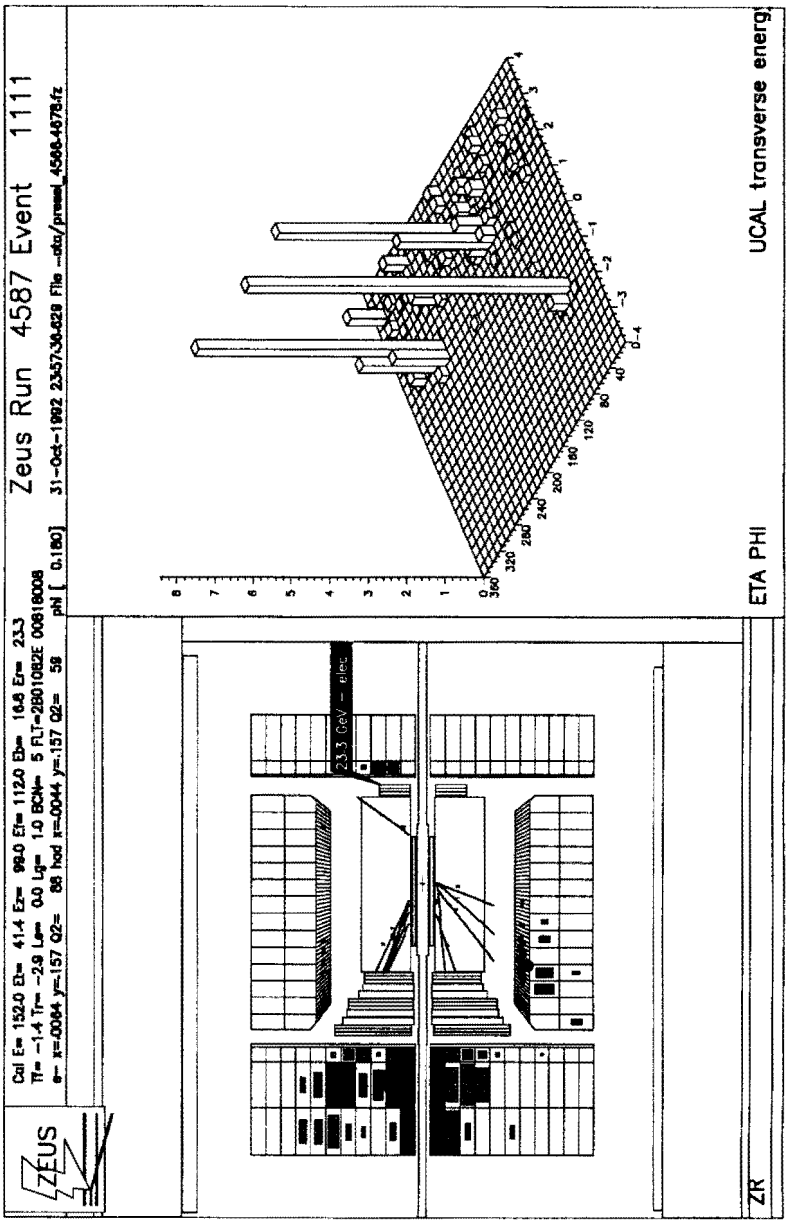

Figure 10.7: *A two jet event.*

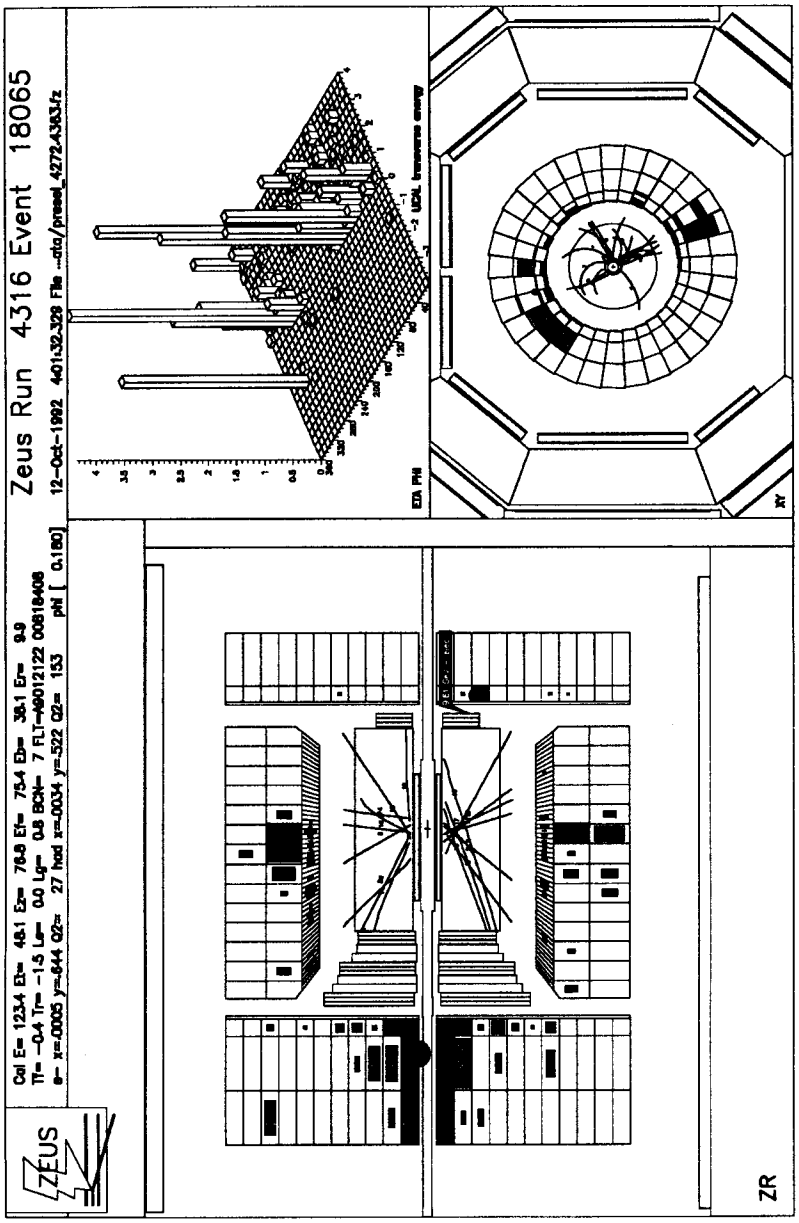

Figure 10.8: *A two jet event.*

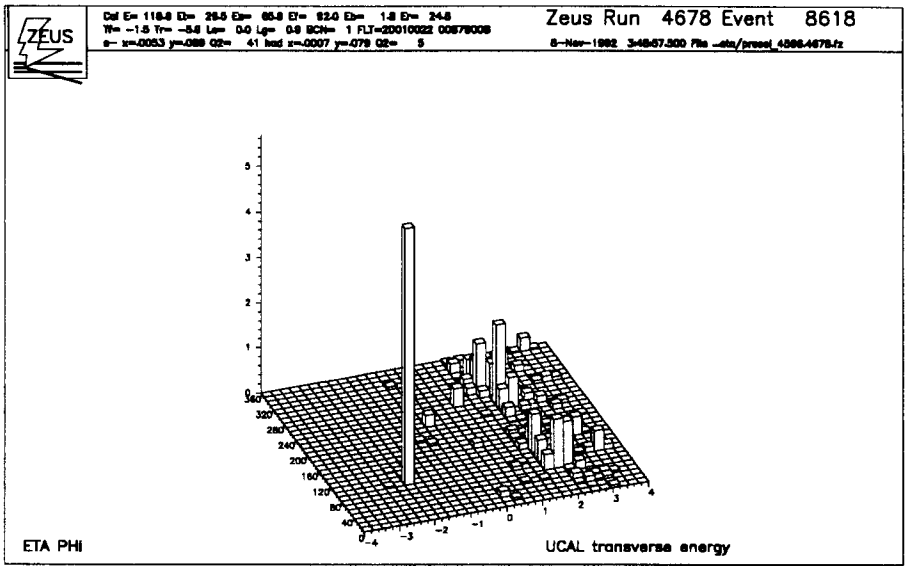

Figure 10.9: *A two jet event.*

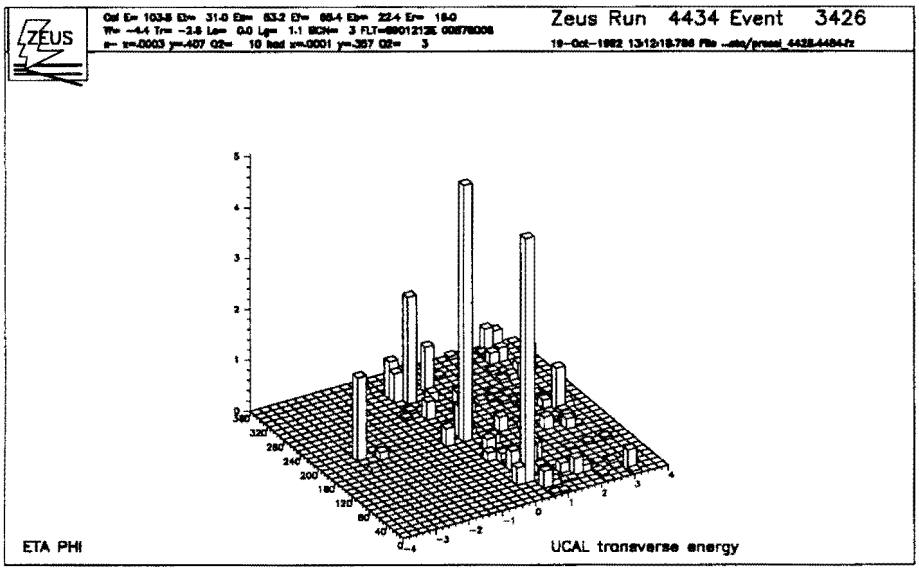

Figure 10.10: *A three jet event.*

References

[1] R. G. Roberts, *The Structure of the Proton*, Cambridge University Press 1990

[2] S. R. Mishra and F. Sciulli, Annu. Rev. Nucl. Part. Sci. **39** (1989) 259

[3] T. Sloan et al, Phys. Rep. **162** (1988) 45

[4] M. Diemoz et al, Phys. Rep. **130** (1986) 293

[5] Proceedings of the nineteenth SLAC Summer Institute on Particle Physics (1991), SLAC-Report-398

[6] R. E. Taylor, H. W. Kendall and J. I. Friedman, Rev. Mod. Phys. **63** (1991) 573-629

[7] J. D. Bjorken, Phys. Rev. **179** (1969) 1547

[8] R. P. Feynman, Phys. Rev. Lett. **23** (1969) 1415
J. D. Bjorken and E. A. Paschos, Phys. Rev. **185** (1969) 1975

[9] C. G. Callan and D. J. Gross, Phys. Rev. Lett. **22** (1969) 156

[10] G. Altarelli and G. Parisi, Nucl. Phys. B126 (1977) 298

[11] M. Virchaux and A. Milsztajn, Phys. Lett. B **274** (1992) 221

[12] H. Plothow-Besch, PDFLIB, Proceedings of the 3rd Workshop on Detector and Event Simulation in High Energy Physics, Amsterdam 1991.

[13] J. F. Owens and W. K. Tung, Annu. Rev. Nucl. Part. Sci. **42** (1992) 291

[14] J. G. Morfin and W. K. Tung, Z. Phys. C **52** (1991) 13

[15] A. D. Martin et al., Durham/RAL preprint, DTP-92-16 and RAL-92-021, Phys. Rev. D **47** (1993) 867

[16] *Physics at HERA*, Proceedings of the HERA physics workshop, Hamburg 1992

[17] G. Ingelman, LEPTO, Proc. of the Workshop on Physics at HERA vol. 3 (1992) 1366

[18] H. Spiesberger et al., Proc. of the Workshop on Physics at HERA vol. 2 (1992) 798

[19] A. Kwiatkowski et al., Z. Phys. C **50** (1991) 165

[20] A. Kwiatkowski et al., HERACLES, Proc. of the Workshop on Physics at HERA vol. 3 (1992) 1294

[21] M. Derrick et al. (ZEUS collaboration), Phys. Lett. B **293** (1992) 465

[22] W. Buchmüller et al., Proc. of the Workshop on Physics at HERA vol. 2 (1992) 919

[23] M. Derrick et al. (ZEUS collaboration), *Search for Leptoquarks with the ZEUS Detector*, DESY 93-017 (1993), Phys. Lett. B **306** (1993) 173

[24] S. Bentvelsen et al., Proc. of the Workshop on Physics at HERA vol. 1 (1992) 23

[25] J. Feltesse, Proc. of the 1987 HERA Workshop vol. 1 (1987) 33

[26] F. Jacquet and A. Blondel, Proceedings of the study for an *ep* facility for Europe, DESY 79/48 (1979) 391

[27] Proceedings of the Study for an *ep* Facility for Europe, DESY 79/48, Hamburg 1979
HERA, A Proposal for a large Electron–Proton Colliding Beam Facility at DESY, DESY HERA 81/10, Hamburg 1981
R. Brinkman, DESY HERA 88/03
B. H. Wiik, Proc. of the Workshop on Physics at HERA vol. 1 (1992) 1

[28] D. Degele, Proc. Third Eur. Part. Acc. Conf, Berlin March 1992

[29] M. Düren and K. Rith, Proc. of the Workshop on Physics at HERA vol. 1 (1992) 427
HERMES, K. Coulter et al., DESY/PRC 90/1 (1990)

[30] H. Bethe and W. Heitler, Proc. Roy. Soc. **A146** (1934) 83

[31] M. van der Horst, Thesis, University of Amsterdam, 1990

[32] J. Straver, Thesis, University of Amsterdam, 1991

[33] E. Longo, Proc. of the Workshop Experimentation at HERA, Amsterdam (1983) 285

[34] The ZEUS Detector, Technical Proposal, Hamburg March 1986

[35] The ZEUS Detector, Status Report 1993, DESY
The ZEUS Detector, Status Report 1989, DESY

[36] U. Amaldi, Physica Scripta **23** (1981) 409
C. W. Fabjan, *Calorimetry in High Energy Physics*, CERN-EP/85-54 (1985)
C. W. Fabjan and R. Wigmans, *Energy Measurement of Elementary Particles*, Rep. Prog. Phys. **52** (1989) 1519
H. Brückmann et al., *On the Theoretical Understanding of Sampling Calorimeters*, DESY 87-064 (1987)
H. Brückmann et al., Nucl. Instr. and Meth. **A263** (1988) 136
R. Wigmans, *Energy Loss of Particles in Dense Matter, Calorimetry*, NIKHEF-H/87-12 (1987)
R. Wigmans, Nucl. Instr. and Meth. **A259** (1987) 389
R. Wigmans, Annu. Rev. Nucl. Part. Sci. **41** (1991) 133

[37] B. Rossi, *High Energy Particles*, Prentice Hall, New York (1952)

[38] Particle Data Group, *Review of Particle Properties*, Phys. Rev. D **45**, Part 2 (1992) S1

[39] J. Krüger, Habilitationsschrift, DESY F35-92-02 (1992), and references therein

[40] G. Drews et al., Nucl. Instr. and Meth. **A290** (1990) 335

[41] A. Andresen et al. (ZEUS FCAL/RCAL group), Nucl. Instr. and Meth. **A309** (1991) 101

[42] M. Derrick et al. (ZEUS BCAL group), Nucl. Instr. and Meth. **A309** (1991) 77

[43] A. Caldwell et al., Nucl. Instr. and Meth. **A321** (1992) 356

[44] A. Caldwell et al., internal ZEUS note 93-021 (1993)

[45] H. Uijterwaal, Thesis, University of Amsterdam, 1992

[46] H. van der Lugt, Thesis, University of Amsterdam, 1993

[47] A. Bernstein et al., internal ZEUS note 91-114 (1991)

[48] U. Behrens et al., Nucl. Instr. and Meth. **A289** (1990) 115

[49] N. Pavel, internal ZEUS note 91-094 (1991)

[50] G. Hartner and Y. Iga, *the ZEUS detector simulation program MOZART*

[51] A. Dake, GAZE, internal ZEUS note 92-006 (1992)

[52] W. Dorth, LAZE, internal ZEUS note 91-069 (1991)

[53] S. M. Fisher and P. Palazzi, *The ADAMO Data System*, Programmers Manual – Version 3.2 (1992)

[54] R. Brun et al., GEANT, CERN DD/EE/84-1 (1984)

[55] R. Brun and J. Zoll, ZEBRA, CERN Program Library Q100 (1989)

[56] E. Tscheslog, internal ZEUS note 91-038 (1991)

[57] J. C. Hart, internal ZEUS note 92-102 (1992)

[58] R. Glaeser, internal ZEUS note 91-103 (1991)

[59] E. Yourdon, *Modern Structured Analysis*, Prentice Hall 1989
T. de Marco, *Structured Analysis and System Specifications*, Prentice Hall

[60] TEAMWORK, CADRE Technologies Inc., Providence, R.I., USA

[61] P. P. Chen, ACM Transactions on Database Systems vol. 1 no. 1 (1976) 9

[62] H. Kowalski et al., internal ZEUS note 89-049 (1989)

[63] A. Bassi et al., internal ZEUS note 88-097 (1988)

[64] F. Anselmo et al., Particle World, Vol. 2, No. 4 (1991) 108

[65] B. Bambah et al., CERN-TH.5466/89 (1989)

[66] G. Hartner, internal ZEUS note 88-049 (1988)

[67] H. Abramowicz et al., internal ZEUS note 90-088 (1990)

[68] P. de Jong, internal ZEUS note 92-019 (1992)

[69] S. de Jong, Thesis, University of Amsterdam, 1990

[70] W. Bartel et al. (JADE collaboration), Z. Phys. C **33** (1986) 23

[71] B. Flaugher and K. Meier, *A Compilation of Jet Finding Algorithms*, FERMILAB-Conf-90/248-E (1990)

[72] S. Bethke, J. Phys. G: Nucl. Part. Phys. **17** (1991) 1455

[73] S. Catani et al., Phys. Lett. B **269** (1991) 432

[74] T. Sjöstrand, the JETSET 7 program and manual
T. Sjöstrand, Computer Phys. Comm. **39** (1986) 347
T. Sjöstrand and M. Bengtsson, Computer Phys. Comm. **43** (1987) 367

[75] T. Sjöstrand, Computer Phys. Comm. **28** (1983) 229

[76] J. Huth et al., FERMILAB-Conf-90/249-E (1990)

[77] J. Huth, FERMILAB-Conf-89/117-E (1990)
E. Buckley-Geer, FERMILAB-Conf-91/193-E (1991)

[78] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York

[79] P. de Jong, internal ZEUS note 89-074 (1989)

[80] M. Crombie, private communication.

[81] M. Fleischer et al., Proc. of the Workshop on Physics at HERA vol. 1 (1992) 303

[82] J. Hartmann, private communication.

[83] G. Grindhammer, Proc. of the Workshop on Physics at HERA vol. 3 (1992) 1153

[84] A. Petersen et al., Phys. Rev. D **37** (1988) 1
M. Akrawy et al. (OPAL collaboration), Z. Phys. C **47** (1990) 505

[85] M. Arneodo et al. (EMC collaboration), Z. Phys. C **36** (1987) 527

[86] M. Bengtsson et al., Nucl. Phys. B **301** (1988) 554

[87] M. R. Adams et al. (E665 Collaboration), Phys. Rev. Lett. 69 (1992) 1026

[88] G. Ingelman et al., Proc. of the Workshop on Physics at HERA vol. 1 (1992) 255

[89] M. Bengtsson et al., Proc. of the HERA Workshop, Hamburg 1987, vol. 1 (1987) 149
M. Bengtsson and T. Sjöstrand, Z. Phys. C **37** (1988) 465

[90] G. Gustafson, Phys. Lett. B **175** (1986) 453
G. Gustafson and U. Pettersson, Nucl. Phys. B **306** (1988) 746
B. Andersson et al., Z. Phys. C **43** (1989) 621
B. Andersson et al., Nucl. Phys. B **339** (1990) 393

[91] N. Magnussen et al., Proc of the Workshop on Physics at HERA vol. 3 (1992) 1167

[92] L. Lönnblad, ARIADNE, Proc. of the Workshop on Physics at HERA vol. 3 (1992) 1440

[93] T. Sjöstrand, PYTHIA at HERA, Proc. of the Workshop on Physics at HERA vol. 3 (1992) 1405

[94] G. Wolf, *First Results from HERA*, DESY 92-190 (1992)

[95] S. Schlenstedt, private communication.

[96] M. Derrick et al. (ZEUS collaboration), Phys. Lett. B **303** (1993) 183

[97] ZEUS collaboration, *Hadronic Energy Distributions in Deep-Inelastic Electron-Proton Scattering*, DESY 93-068 (1993), to be published in Z. Phys. C.

[98] A. D. Martin et al., Phys. Lett. B **306** (1993) 145

[99] H. Abramowicz, *the* REWEIGHT *function*

[100] A. Levy, private communication.

[101] M. Derrick et al. (ZEUS collaboration), Phys. Lett. B **297** (1992) 404

[102] ZEUS collaboration, *Observation of two-jet production in deep inelastic scattering at HERA*, DESY 93-030 (1993), Phys. Lett. B **306** (1993) 158

Summary

In 1992, the electron-proton collider HERA has produced its first luminosity for the experiments H1 and ZEUS, hereby continuing a long and fruitful tradition of deep inelastic lepton-hadron scattering by opening up a new kinematical region.

In two periods, the ZEUS detector has recorded data corresponding to a total integrated luminosity of 27.9 nb^{-1}. The ZEUS detector was not complete in 1992; the data has been taken with a modified trigger, and in the analysis the high resolution uranium calorimeter, the beam pipe counters and the luminosity monitor have played the most important role; the data from the inner tracking chambers has only been used to reconstruct, if possible, the z coordinate of the interaction vertex point. In general, the detector has operated very well.

Although the fact that the delivered HERA luminosity in 1992 was a factor 1000 below the design luminosity prohibits the stringent QCD test of the evolution of structure functions at high Q^2, HERA has enabled many new physics opportunities by opening the new kinematical region at low x. In this region, perturbative QCD does not make rigorous predictions, and a variety of predictions on the behaviour of structure functions exist, each with different consequences. The HERA data has also enabled the testing of QCD–inspired models on fragmentation and multiple gluon radiation.

In this thesis, it is shown that the hadronic energy flow (resulting from the struck proton) in deep inelastic scattering events can be measured with a high resolution calorimeter, and used to distinguish between a number of these models, test the predictions of various structure function parametrizations in the new kinematical region, and find evidence for the occurring of higher order processes involving gluons in deep inelastic scattering.

In the first chapters, the kinematics and physics at HERA, the collider itself, and the ZEUS detector are described. Hereby the role of calorimetry, i.e. the measurement of particle energies, is emphasized. In particular the design, layout, readout, and calibration of the ZEUS calorimeter is treated in some detail. These chapters on HERA and ZEUS thus described the hardware with which the measurements are performed.

In chapter 5 and 6, the off-line software is described that is needed to perform the measurements. In chapter 5, the general environment of the ZEUS off-line software,

and some parts of the detector and trigger simulation and the event reconstruction program, are described. Chapter 6 treats in detail the reconstruction software for the uranium calorimeter, which has been designed and implemented by the author of this thesis. It is shown how the geometry has been set up, how the energy in the cells is obtained, and how local and global clusters are found. Local clusters, or condensates, are intended to describe showers of individual particles; global clusters are jet-like objects. It is shown how these objects are identified, and the efficiency of the reconstruction procedure is presented.

Chapter 7 describes a number of Monte Carlo generator programs, and the fragmentation and QCD cascade models they implement.

The application to ZEUS data is described in chapters 8 to 10. The presented analysis has been carried out by the author of this thesis as a parallel and independent second analysis to a similar analysis carried out at DESY, which has resulted in two published ZEUS papers [97, 102]. In chapter 8, the status of HERA and ZEUS in 1992 is described, and the recorded data is introduced. The selection of deep inelastic neutral current events in the 1992 data sample is described in detail. We find, following the described criteria, 3449 events at $Q^2 > 10$ GeV^2, which corresponds to an observed inclusive cross section of 130 ± 15 nb. This number corresponds best to the predictions of a structure function parametrization that steeply rises at low x, such as MRSD-, and is significantly higher than a flat parametrization like MTB1.

In chapter 9, various distributions of the reconstructed kinematics and the hadronic energy flow are presented and compared to Monte Carlo data. In general, these distributions are well described by the generator programs LEPTO (ME+PS) and ARIADNE. The distributions that are sensitive to the structure function parametrization are, again, in general best described by MRSD-, whereas MTB1 agrees the least with the data. The models which include only the lowest order DIS matrix elements plus a parton shower approach (for higher order effects) with scale W^2 or Q^2 both fail to describe the data: the W^2 scale predicts too much radiation, the Q^2 scale too little. The differences between the two models can be understood from the fact that the 1992 sample is characterized by a low Q^2, but a high W^2. The used versions of LEPTO and ARIADNE do describe the right amount of gluon radiation, but do not fully describe the data in the very forward (target remnant) region.

In chapter 10, a cone jet finding algorithm, introduced in chapter 6, is applied to the data. Jets are found with properties that are well described by the Monte Carlo. Events with two or more jets outside the beam pipe region have been found in the DIS sample. It is shown that for events with two jets, the two jets are generally back-to-back in azimuth ϕ, and populate the regions of high W^2 and E_T, and low x. These events can be interpreted as being caused by higher order processes in deep inelastic scattering in which gluons play an explicit role, such as boson-gluon fusion, and QCD-Compton scattering.

Samenvatting

In de *hoge-energie fysica*, ook wel *elementaire deeltjes fysica* genoemd, zijn we geinteresseerd in de fundamentele bouwstenen van de materie, en in de wisselwerkingen tussen deze bouwstenen die maken dat stabiele vormen van materie kunnen bestaan.

Stoffen bestaan uit molekulen, en deze weer uit atomen. Er bestaan slechts 92 verschillende soorten atomen in de vrije natuur, en deze bestaan op hun beurt weer uit een elektrisch positief geladen *kern*, omringd door een wolk van negatief geladen *elektronen*, voor ieder soort atoom een ander aantal. In de atoomkern kunnen we ook weer nieuwe deeltjes onderscheiden: de elektrisch positief geladen *protonen*, en de heel sterk op protonen lijkende en ongeveer even zware, maar ongeladen, *neutronen*. Het meest eenvoudige atoom, waterstof, heeft 1 proton als kern, en 1 elektron daaromheen. Zuurstof heeft 8 protonen, 8 neutronen, en 8 elektronen. Bijna de gehele massa van het atoom zit in de kern, terwijl het gedrag van de elektronen belangrijk is voor de chemische reacties van stoffen. Experimenten in de hoge-energie fysica houden zich bezig met het kijken naar objecten ter grootte van, of kleiner dan, het proton, dat een afmeting van ongeveer 10^{-13} cm heeft. Dat is, momenteel, alleen maar mogelijk met enkele van de allergrootste wetenschappelijke instrumenten die bestaan: de deeltjesversnellers. Deze creëren bundels, waarin groepen van deeltjes worden versneld tot bijna de lichtsnelheid, en waarin de afzonderlijke deeltjes een enorme energie verkrijgen. Men laat deze deeltjes botsen op een stilstaand doelwit (*target*), of men laat twee bundels deeltjes frontaal tegen elkaar botsen; het resultaat van zulke botsingen wordt gemeten met grote en gecompliceerde detectoren die de ogen van de natuurkundige vormen. Uit de metingen kan teruggerekend worden hoe de botsingen zijn verlopen, en daaruit kan weer informatie over de structuur van de deeltjes en hun wisselwerkingen gewonnen worden.

Het is gebleken, rond 1970, dat ook het proton en het neutron weer bestaan uit nog kleinere deeltjes, de *quarks*. Zo bestaat het proton in wezen uit twee quarks van het type *up*, en één quark van het type *down*, en het neutron uit één quark van het type *up*, en twee quarks van het type *down*. Voor het elektron heeft men zo'n structuur (nog?) niet kunnen vinden. Het *up* quark, het *down* quark, en het elektron, zijn stabiel en vallen niet uiteen in andere deeltjes. Al ver voor 1970 was ontdekt dat in botsingen bij hoge energieën nog veel meer, zwaardere, deeltjes

geproduceerd konden worden. Binnen het quark–model kunnen al deze deeltjes, lichte zowel als zware, opgebouwd worden uit een totaal van zes types quarks, of maken deel uit van de familie van de *leptonen* (per definitie zijn dit deeltjes die niet gevoelig zijn voor de krachten die heersen in de kern van een atoom), die bestaat uit het elektron en twee zwaardere varianten daarop, en drie tamelijk spookachtige deeltjes genaamd neutrino's. De zwaardere deeltjes zijn niet stabiel, en vallen uiteindelijk weer uiteen in deeltjes met slechts *up* en *down* quarks, of in elektronen en neutrino's.

Vier soorten krachten zijn verantwoordelijk voor alle wisselwerkingen tussen elementaire deeltjes. De zwaartekracht werkt op deeltjes met massa, en is belangrijk voor ons begrip van het heelal. Elektromagnetische krachten spelen een rol in zowel elektrische als magnetische verschijnselen, werken op deeltjes met elektrische lading, en zijn belangrijk voor de bouw van kristallen, molekulen en atomen, en dus veelal voor de verschijnselen die een rol spelen in het normale leven. De sterke kernkracht werkt op quarks en is verantwoordelijk voor de stabiliteit en bouw van protonen, neutronen en de atoomkern. De zwakke kernkracht, tenslotte, werkt op alle deeltjes, is zwak, en vaak verantwoordelijk voor verval van instabiele deeltjes zoals deze voorkomen in radioactieve stoffen.

Deze vier krachten zijn niet even sterk, en hun sterkte hangt af van de energieën van de deeltjes waartussen de krachten werken. Bij energieën die typisch zijn voor de huidige deeltjesversnellers, is de sterke kernkracht 10^3 maal sterker dan de elektromagnetische kracht, 10^{12} maal sterker dan de zwakke kernkracht, en 10^{43} maal zwakker dan de zwaartekracht. De zwaartekracht speelt dus voor ons geen belangrijke rol in de interacties tussen elementaire deeltjes.

De wisselwerkingen tussen elementaire deeltjes, zoals deze gemeten zijn in vroegere en tegenwoordige experimenten bij deeltjesversnellers, worden heel goed beschreven door een verzameling theorieën die samen het *standaard model* genoemd worden. Het standaard model omvat beschrijvingen van elektromagnetisme, de sterke kernkracht en de zwakke kernkracht, die zowel in overeenstemming zijn met de speciale relativiteitstheorie als met de quantummechanica. De speciale relativiteitstheorie beschrijft verschijnselen die zich voordoen wanneer deeltjes een snelheid dicht bij die van het licht bereiken, zoals in versnellers gebruikelijk is. De quantummechanica is nodig voor een beschrijving van het gedrag en de wisselwerkingen van verzamelingen van weinig deeltjes, die heel anders is dan het bekende gedrag van grote objecten. Volgens de quantummechanica kan het gedrag van individuele deeltjes niet langer exact deterministisch voorspeld worden, maar speelt het toeval een belangrijke rol in interacties tussen elementaire deeltjes. Theoretische berekeningen kunnen wisselwerkingen tussen twee deeltjes dus niet meer van geval tot geval exact beschrijven, maar alleen statistisch, d.w.z. dat ze de kans bepalen dat twee deeltjes op een bepaalde manier zullen wisselwerken. De interactie tussen deze deeltjes, en dus de kracht die de deeltjes van elkaar ondervinden, kan beschreven worden alsof de deeltjes onderling voortdurend weer andere, zeer specifieke, deeltjes uitwisselen: *fotonen* voor de elektromagnetische kracht, *gluonen* voor de sterke kernkracht, en de zogenaamde W and Z *bosonen* voor de zwakke kernkracht. De theorie voor de sterke

kernkracht wordt *quantumchromodynamica* (QCD) genoemd, zij beschrijft precies hoe quarks elkaars aanwezigheid voelen door onderling gluonen uit te wisselen.

Ondanks het succes van het standaard model geeft het waarschijnlijk geen afdoende beschrijving van de wereld van elementaire deeltjes. Er zijn goede aanwijzingen dat bij zeer hoge energieën de afzonderlijke krachten zoals we die nu voelen samensmelten tot één kracht. Het standaard model beschrijft hoe dit gebeurt voor de elektromagnetische en de zwakke kernkracht, maar niet voor de sterke kernkracht. Een consistente quantummechanische beschrijving van de zwaartekracht, en daarmee een opname van de zwaartekracht in het standaard model, is tot nu toe onmogelijk gebleken. Ook zijn er een groot aantal vrije parameters in het standaard model, waarvan de waarde's niet voorspeld worden, maar die heel belangrijk zijn voor de consistentie van de theorie.

Een belangrijke beperking van het gebruik van de theorieën in het huidige standaard model is het feit dat de manier waarop grootheden, zoals de kans op een interactie, berekend worden zodanig is dat het gebruik op problemen stuit in situaties waarin de krachten zeer sterk zijn. Nu is het zo dat energieën (van deeltjes) en afmetingen (afstanden tussen deeltjes) aan elkaar zijn gerelateerd: wanneer twee deeltjes een hoge energie hebben kunnen zij elkaar veel dichter naderen dan wanneer zij een lage energie hebben. Voor QCD geldt dat de sterke kernkracht zwak is bij zeer hoge energieën (overeenkomend met korte afstanden tussen quarks, afstanden veel kleiner dan de afmetingen van een proton), maar vrij sterk bij lagere energieën (bijvoorbeeld overeenkomend met afstanden vergelijkbaar met de afmetingen van een proton). De sterke kernkracht tussen twee quarks kan vergeleken worden met die van een veer: wanneer de quarks vlak bij elkaar zijn is de kracht klein, maar als de quarks ver van elkaar zijn is de kracht groot. Dit wil zeggen dat het binnen QCD vrij gemakkelijk is om te berekenen hoe twee quarks met zeer hoge energieën wisselwerken, maar dat het uitermate moeilijk is om precies te berekenen hoe quarks een deeltje zoals een proton vormen, hetgeen toch ook een interessant probleem is.

In het licht van deze problemen van het standaard model is het dan ook niet verwonderlijk dat de huidige experimenten in de hoge-energie fysica veelal een drietal doelen hebben: hetzij vormen van wisselwerkingen te vinden die in tegenspraak zijn met het standaard model, hetzij nieuwe deeltjes te vinden die geen plaats hebben in het standaard model, hetzij wisselwerkingen te onderzoeken onder zodanige omstandigheden dat deze binnen het standaard model niet goed beschreven kunnen worden. Tot nu (juli 1993) toe heeft nog geen enkel experiment een duidelijke tegenspraak met het standaard model gevonden, en ook lijken alle gevonden deeltjes in het model van quarks en leptonen te passen.

Het onderzoek dat in dit proefschrift wordt beschreven maakt deel uit van het onderzoekprogramma naar de structuur van het proton. Hoewel dit programma alle drie hierboven genoemde doelen omvat, beslaat het onderzoek in dit proefschrift toch vooral het derde doel, waarbij met name een aantal modellen van de structuur van het proton, en van QCD-interacties bij energieën die overeenkomen met afstanden in de orde van grootte van de afmetingen van een proton, worden onderzocht. In het vervolg van deze samenvatting zal worden uitgelegd hoe dit is

gedaan.

In 1992 is bij het DESY laboratorium in Hamburg een nieuwe versneller, HERA, in gebruik genomen. In een ringvormige tunnel met een omtrek van ruim 6300 meter, 20 tot 40 meter onder de grond gelegen, worden in tegenovergestelde richtingen bundels elektronen en protonen versneld, en op twee plaatsen frontaal tegen elkaar geschoten. Deze twee punten zijn omringd door twee grote en gecompliceerde detectoren, H1 en ZEUS geheten. Wanneer HERA op volle toeren draait, zijn de elektronen en de protonen gegroepeerd in 200 pakketjes van enkele tientallen miljarden deeltjes elk, die elkaar 10 miljoen maal per seconde passeren. De deeltjes zijn echter dermate klein dat ze elkaar tijdens zulke passages meestal met rust laten, maar af en toe, ongeveer 1 maal per seconde per pakketje (en voor de echt interessante gebeurtenissen ongeveer 1 maal per minuut per pakketje), botsen een elektron uit de ene bundel en een proton uit de andere op elkaar. Daarbij wordt het elektron verstrooid (het verandert van richting en krijgt meestal een andere energie), en wordt het proton veelal in stukken geslagen. De individuele quarks uit het proton herschikken zich, en er kunnen nieuwe quarks worden geproduceerd; de vrijkomende quarks kunnen niet vrij bestaan, maar vormen, door een veelheid aan quark–gluon en gluon–gluon interacties, een stortvloed van nieuwe deeltjes die vanuit het botsingspunt alle kanten uitvliegen. De twee detectoren rondom het botsingspunt zijn in staat deze deeltjes waar te nemen en hun eigenschappen te meten, en zo kan inzicht verkregen worden in de structuur van het proton. Het geheel kan vergeleken worden met een microscoop: het te bekijken object (een proton) wordt beschenen met een bundel licht (fotonen afkomstig van de bundel elektronen). De afmeting van het kleinste wat nog waarneembaar is met een microscoop wordt bepaald door de golflengte van het gebruikte licht, voor HERA bedraagt deze ongeveer 10^{-14} tot 10^{-16} cm (ook hierin speelt de energie van de elektronen en protonen een rol: hoe hoger de energie, hoe kleiner de kleinste golflengte van het licht). Men verwacht met HERA voor het eerst in staat te zijn om voorspellingen van QCD te testen bij dergelijke zeer kleine afstanden, en eventuele substructuur in quarks of elektronen te kunnen ontdekken.

De in dit proefschrift beschreven metingen zijn de eerste verkregen met de ZEUS detector gedurende 1992. De hoeveelheid data genomen in 1992 is veel te gering voor een volledige uitbuiting van de mogelijkheden bij HERA, maar er zijn al wel reeds interessante nieuwe resultaten verkregen.

In hoofdstuk 2 van dit proefschrift worden de natuurkundige mogelijkheden van HERA beschreven. Het is gebruikelijk de structuur van het proton te beschrijven in de vorm van een aantal *structuurfuncties*. Binnen de QCD theorie kunnen deze geinterpreteerd worden als het resultaat van de eigenschappen van de quarks en gluonen waaruit het proton bestaat. Eén van de structuurfuncties, bijvoorbeeld, beschrijft dan welke fractie x (x ligt tussen 0 en 1) van de impuls van het proton wordt gedragen door een bepaald quark of gluon in dat proton. Het blijkt dat in QCD deze functies ook afhangen van de resolutie, of nauwkeurigheid, waarmee het proton bekeken wordt (d.w.z. van de golflengte van het gebruikte licht), en QCD voorspelt precies hoe. Er bestaat al data van vroegere experimenten over deze

structuurfuncties, maar HERA zal veel 'dieper' in het proton kunnen kijken dan deze eerdere experimenten, en dus veel nauwkeuriger tests van QCD uitvoeren, en eventueel geheel nieuwe ontdekkingen doen. In hoofdstuk 2 wordt ook beschreven hoe uit de metingen van een gedetecteerde botsing de variabele x, en de resolutie parameter Q^2 (gerelateerd aan de golflengte van het gebruikte licht: hoe groter Q^2, hoe kleiner de golflengte van het gebruikte licht) kunnen worden berekend.

In hoofdstuk 3 wordt de HERA versneller zelf beschreven, hoe men kan bepalen hoeveel elektronen en protonen door HERA op elkaar zijn afgevuurd, en welke processen de metingen kunnen verstoren (de zogenaamde *achtergrond*). HERA levert namelijk niet alleen elektron–proton botsingen, maar ook botsingen van elektronen en protonen met gasmolekulen die in de versnellerbuis zijn achtergebleven, en die ook vaak door de detectoren gemeten worden. Dit type achtergrond botsingen vindt vele duizenden malen per seconde plaats, dus veel en veel vaker dan de gezochte elektron–proton botsingen. Elke detector bevat dan ook zeer uitgebreide en complexe systemen, die samen de *trigger* worden genoemd, om de interessante botsingen te selecteren uit de achtergrond.

De ZEUS detector is het onderwerp van hoofdstuk 4. Figuur 4.1 geeft een doorsnede van de detector langs de as van de bundels: de elektronen komen van links en de protonen van rechts. Figuur 4.2 geeft een doorsnede loodrecht op de bundel as. Vrijwel alle metingen in dit proefschrift zijn gemaakt met een zeer belangrijk onderdeel van de ZEUS detector: de *calorimeter*. De principes van de werking van dit apparaat, en de structuur van de ZEUS calorimeter worden ook uitvoerig beschreven in dit hoofdstuk. Samenvattend kan worden gesteld dat de calorimeter de energie van in de botsing geproduceerde deeltjes meet door hen te absorberen in een grote hoeveelheid zwaar materiaal (uranium), waarbij een fractie van de energie in een soort plexiglas (*scintillator*) wordt geabsorbeerd en omgezet in een lichtsignaal. De calorimeter bestaat uit een groot aantal lagen, en iedere laag bevat om en om een plaat uranium en een plaat scintillator. Het licht van een aantal scintillator lagen wordt verzameld, en geleid naar een detector die de hoeveelheid licht kan meten, hetgeen een maat is voor de energie van het binnenvallende deeltje. In totaal is de calorimeter verdeeld in 5918 verschillende cellen (ieder bestaande uit tussen de 23 en 80 lagen en gelegen in een ander deel van de calorimeter); een grote hoeveelheid elektronika zet de waarden van de gemeten hoeveelheden licht in elk van de cellen en voor iedere mogelijke botsing om in een getal, dat vervolgens bewaard wordt en, als de botsing geaccepteerd wordt door de trigger, weggeschreven naar een magneetband.

De ruwe informatie op de magneetbanden is niet direkt geschikt voor een analyse van de resultaten. Eerst wordt de data verder verwerkt door een speciaal ontwikkeld programma, het zogenaamde *reconstructie* programma, dat de ruwe signalen kan interpreteren en omzetten naar grootheden die de identiteit, positie en energie van de geproduceerde deeltjes weergeven. Zoals de naam van dit programma al suggereert, probeert het de botsing te reconstrueren, d.w.z. terug te rekenen welke deeltjes in de botsing zijn vrijgekomen, en met welke eigenschappen. De verschillende programma's die ontwikkeld zijn binnen de ZEUS collaboratie, en de omgeving waarin

zij werken, zijn beschreven in hoofdstuk 5. Hoofdstuk 6 gaat in detail in op het reconstructie programma voor de calorimeter, dat door de auteur van dit proefschrift is ontworpen, geschreven en getest.

In het calorimeter reconstructie programma worden gegevens betreffende de structuur van de calorimeter, en de afmetingen en lokatie van de calorimeter cellen, gebruikt om de ruwe signalen op de magneetbanden te interpreteren. Daarbij wordt geprobeerd om de ruis (toevallige signalen die niets met invallende deeltjes te maken hebben) zoveel mogelijk te onderdrukken, en een zo nauwkeurig mogelijke schatting van de werkelijk gedeponeerde energie in de cellen te maken. Vervolgens kunnen de cellen op een aantal verschillende manieren worden gegroepeerd. Het doel van deze samenvoeging van cellen is het sommeren van de energie van cellen die om de een of andere reden bij elkaar horen; cellen kunnen bij elkaar horen wanneer hun energie afkomstig is van één deeltje dat door de calorimeter geabsorbeerd wordt (dan heet de groep cellen een *condensate*), of van een groep deeltjes die op hun beurt afkomstig zijn van één quark dat in de elektron–proton botsing is vrijgekomen (dan heet de groep cellen een *cluster* of *jet*). Uit de verdeling van de energie in de cellen die horen bij een groep kan veelal bepaald worden welk soort deeltjes door de calorimeter gemeten zijn. Het juist meten van elektronen is bijzonder belangrijk, omdat deze deeltjes vaak het signaal zijn dat er een interessante elektron–proton botsing heeft plaatsgevonden.

Een belangrijk hulpmiddel voor het uitvoeren van experimenten in de hoge-energie fysica bestaat uit computer programma's die botsingen tussen deeltjes en de respons daarop van de detector kunnen simuleren. Vanwege de belangrijke rol die het toeval in dit soort processen speelt worden in de simulatie ervan technieken gebruikt die *Monte Carlo* technieken worden genoemd, en de programma's heten dan ook vaak zelf Monte Carlo's. Deze programma's worden gebruikt om informatie te verkrijgen over dingen waarvan het veel te gecompliceerd of zelfs onmogelijk is om ze exact te berekenen. Dit omvat onder meer de efficiency van de detector en de elektronica om goede botsingen te meten en te selecteren, en de mate waarin toch nog achtergrond processen geselecteerd worden [2]. Met behulp van Monte Carlo's worden de meest optimale selectie criteria gezocht. Ook de effectiviteit van het reconstructie programma kan worden onderzocht met een Monte Carlo; voor de calorimeter is dit beschreven in hoofdstuk 6. Tevens kan de nauwkeurigheid waarmee x en Q^2 gereconstrueerd kunnen worden worden berekend.

In hoofdstuk 7 worden een aantal van deze simulatie programma's beschreven. In deze programma's zijn de theorieën uit het standaard model verwerkt, maar dit is niet voldoende. Zoals eerder beschreven kunnen niet alle QCD processen die zich afspelen wanneer quarks of gluonen vrijkomen uit een proton en zelf nieuwe deeltjes produceren exact worden berekend. De Monte Carlo's bevatten dus modellen voor zulke processen, en er zijn een aantal verschillende modellen hiervoor op de markt. Door de gemeten botsingen te vergelijken met de gesimuleerde botsingen, kan infor-

[2] In de ontwerpfase van een experiment worden zelfs vaak de eisen waaraan de detector zelf moet voldoen onderzocht met Monte Carlo programma's.

matie over de juistheid van de modellen en de achterliggende gedachten gewonnen worden. Zo'n vergelijking wordt ook in dit proefschrift gemaakt.

Een Monte Carlo programma heeft ook informatie nodig over de structuurfuncties van het proton. Zoals eerder vermeld zijn deze gemeten door vroegere experimenten, maar niet onder de omstandigheden zoals deze zich voordoen bij HERA. Het gedrag van deze structuurfuncties bij HERA is dus niet bekend: er bestaan een aantal verschillende modellen die, hoewel ze alle in overeenstemming zijn met de data van de vroegere experimenten, toch onderling verschillen voor HERA. Ook deze modellen kunnen experimenteel onderzocht worden door de gemeten botsingen te vergelijken met de gesimuleerde.

In hoofdstuk 8 wordt beschreven hoe de elektron–proton botsingen waarmee deze analyse is uitgevoerd geselecteerd zijn uit alle data die door ZEUS in 1992 is genomen. ZEUS heeft dat gedaan in twee periode's, één in de zomer en één in de herfst, waarbij in de herfst effectief ongeveer 10 maal zoveel data is genomen als in de zomer. De detector heeft daarbij zeer goed gewerkt, en met name de calorimeter heeft een centrale rol gespeeld in het verzamelen en analyseren van de data. In hoofdstuk 8 wordt precies beschreven welke criteria zijn aangelegd om goede botsingen te scheiden van achtergrond botsingen. De achtergrond in het laatste stadium van de analyse, dus na selectie door de trigger en na toepassing van de meest eenvoudige selectiecriteria, bestaat voornamelijk uit echte elektron–proton botsingen, waarvan het elektron dermate weinig is afgebogen dat het door de detector niet gemeten wordt, maar waarbij abusievelijk een ander deeltje aangezien is voor het verstrooide elektron. Ook deze achtergrond kan echter worden onderdrukt, en uiteindelijk blijven na toepassing van alle selectiecriteria 3449 botsingen over voor verdere analyse. Twee van deze botsingen zijn bij wijze van voorbeeld getoond in figuur 8.6 en figuur 8.7. Deze figuren tonen plaatjes van een zogenaamd *event display* programma, dat schematisch, en op een aantal verschillende manieren, de signalen van de detector laat zien voor een bepaalde elektron–proton botsing. Figuur 8.6 toont een veel voorkomend soort botsing, waarbij het elektron slechts weining wordt afgebogen en vrijwel dezelfde energie behoudt. Figuur 8.7 toont een bijzondere botsing, waarbij het elektron meer dan 90° wordt afgebogen en de resolutieparameter Q^2 zeer hoog is. Ook is fraai te zien hoe een quark uit het proton wordt geslingerd en een aantal nieuwe deeltjes vormt die veel energie deponeren in de calorimeter.

Uit de hoeveelheid geselecteerde botsingen en informatie over het totale aantal elektronen en protonen dat door HERA aan ZEUS is geleverd en die de kans hebben gehad om te botsen, kan de effectieve kans berekend worden op een elektron–proton botsing (technisch uitgedrukt in een zogenaamde werkzame doorsnede). Deze kans is een interessante parameter, omdat hij afhangt van de structuurfunctie, en dus verschilt tussen de verschillende daarvoor gebruikte modellen. De metingen beschreven in dit proefschrift laten zien dat de data genomen met ZEUS in 1992 al onderscheid kunnen maken tussen verschillende structuurfunctie modellen, en tonen aan dat de elektron–proton botsingskans groter is dan de meeste modellen voorspellen, en alleen beschreven kan worden met modellen waarin het proton zeer sterk bevolkt

wordt met quarks en gluonen met een lage fractie van de proton impuls.

In hoofdstuk 9 worden een aantal grootheden, die gebaseerd zijn op de gemeten energie in de calorimeter in de geselecteerde botsingen, berekend en gepresenteerd, en vergeleken met de verwachtingen die volgen uit de verschillende modellen die in de Monte Carlo's verwerkt zijn. De conversiefactor die nodig is om de energie van een deeltje dat door de calorimeter geabsorbeerd is te berekenen uit de gemeten hoeveelheid licht in de calorimeter, is gedurende de afgelopen jaren zorgvuldig bepaald in testopstellingen, en wordt door ZEUS voortdurend gecontroleerd. Met behulp van de data wordt in het begin van hoofdstuk 9 van dit proefschrift geprobeerd te onderzoeken of deze conversiefactor ook nog klopt nu de calorimeter is opgesteld in de ZEUS detector en blootgesteld aan werkelijke botsingen bij HERA, en dit lijkt het geval te zijn. Sommige van de in hoofdstuk 9 gepresenteerde grootheden zijn gevoelig voor het gebruikte structuurfunctie-model, en opnieuw blijkt dat de metingen wijzen op een structuurfunctie die alleen beschreven kan worden met modellen waarin het proton zeer sterk bevolkt wordt met quarks en gluonen met een lage fractie van de proton impuls. Ook wordt aangetoond dat een tweetal modellen betreffende de QCD-processen die zich afspelen bij de relatief lage energieën, wanneer quarks deeltjes vormen, de data goed kunnen beschrijven, en een tweetal andere modellen niet.

In het laatste hoofdstuk, hoofdstuk 10, wordt gezocht naar botsingen waarin de getroffen quarks afzonderlijk duidelijke bundels deeltjes produceren die met de calorimeter kunnen worden gemeten. Zulke deeltjesbundels worden *jets* genoemd, en kunnen worden gereconstrueerd met behulp van de programmatuur die in hoofdstuk 6 is beschreven. Met name interessant zijn de botsingen waarin meer dan één jet wordt gevonden: deze zijn aanwijzingen voor het bestaan van botsingen waarin niet simpelweg één quark wordt getroffen, maar waarin twee of meer quarks of gluonen uit het proton worden weggeschoten. Dit soort processen komen niet voor in het simpelste model van het proton, maar worden wel voorspeld door QCD, en zijn een rechtstreeks bewijs voor het feit dat protonen ook gluonen bevatten. Inderdaad worden zulke botsingen gevonden, met de eigenschappen die ook door de Monte Carlo programma's worden verwacht. De figuren 10.7 tot en met 10.10 laten event display plaatjes van enkele van zulke botsingen zien. Met name in het plaatje met de naam 'UCAL transverse energy' vallen de twee jets goed op: de hoge piek in weinig cellen bij kleine waarden van 'eta' is het verstrooide elektron, de twee andere pieken zijn de jets. Figuur 10.10 laat een botsing zien met zelfs 3 jets, ook dergelijke botsingen worden door QCD voorspeld.

Geconcludeerd kan worden dat de ZEUS detector, en de calorimeter in het bijzonder, in zijn eerste jaar van het nemen van data bij HERA, goed gewerkt heeft. In dit proefschrift is beschreven hoe de ruwe genomen data bewerkt en gebruikt kan worden in een onderzoek naar de structuur van het proton. Daarbij is reeds gebleken dat een aantal modellen wel, en een aantal andere modellen niet in overeenstemming zijn met de data. Het lijkt aannemelijk dat ZEUS, wanneer HERA eenmaal op volle toeren draait, nog vele nieuwe bijdragen aan het onderzoek naar het proton zal kunnen leveren.

Acknowledgements

The work described in this thesis has been carried out within the context of a large collaboration of physicists and technicians, and has benefitted from the ideas and help of a large number of people at NIKHEF and at DESY. I would like to mention the following persons in particular, in the hope to be forgiven by those I may have forgotten.

I would like to thank my promotor, Armin Tenner, for accepting me as a student and introducing me to high energy physics, for his ideas during the initial phase of the design of the calorimeter reconstruction software and the many suggestions that followed later, and for the critical eye with which he has read this manuscript.

I thank Jos Engelen for his continuous support and interest during my time as a PhD student in the ZEUS group at NIKHEF. I greatly appreciate the interest, comments and help of Henk Tiecke, and I also thank Els de Wolf and Paul Kooijman for their suggestions. The many discussions with these staff members at NIKHEF have had a great influence on my knowledge of experimental high energy physics.

I would like to thank Malcolm Derrick for his guidance in the analysis of the data, and I am also very much indebted to Mike Crombie, Nick Brook, Paul Makkar, Steve Magill, Guy Cases, and Fernando Barreiro.

Many people in the ZEUS offline reconstruction group have helped me, and I would like to thank Jutta Hartmann, Gerd Hartner, Yoshihisa Iga, Matthias Kasemann, Reiner Gläser, Evelin Tscheslog, Till Poser, Nikolaj Pavel, Stefan Schlenstedt, Jacek Gajewski, Ken Long, John Hart, Norman McCubbin, and Jim Whitmore.

I am grateful to the whole group that has designed, built, and operated (very succesfully) the ZEUS calorimeter, and in particular to Robert Klanner, Tatsu Tsurugai, and Rik Yoshida.

It has been a great pleasure to cooperate with all the PhD students and post-docs in the ZEUS group at NIKHEF: Sijbrand de Jong, Jan Straver, Henk Uijterwaal, Andrej Dake, Hermen van der Lugt, Stan Bentvelsen, Roel van Woudenberg, Marc de Kamps, Marcel Vreeswijk, Andres Kruse, and Vivian O'Dell. Their contributions have been numerous.

I would like to thank the NIKHEF computer– and user support group, and in particular Eric Wassenaar and Willem van Leeuwen, for the facilities and support that have been available at NIKHEF.

Last, but by no means least, I thank Suzanne Roeffen for the cover design and for much, much more.